

From the Department of Medicine, Clinical Epidemiology Unit  
Karolinska Institutet, Stockholm, Sweden

**LIFESTYLE AND INFECTIONS:  
AN EPIDEMIOLOGICAL APPROACH ON THE ROLE OF  
OBESITY, PHYSICAL ACTIVITY, SLEEP AND STRESS**

Francesca Ghilotti



**Karolinska  
Institutet**

Stockholm 2020

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet

Printed by E-Print AB 2020

Typeset by the author using  $\text{\LaTeX}$  2<sub>ε</sub>

©Francesca Ghilotti, 2020

ISBN 978-91-7831-726-4

# LIFESTYLE AND INFECTIONS: AN EPIDEMIOLOGICAL APPROACH ON THE ROLE OF OBESITY, PHYSICAL ACTIVITY, SLEEP AND STRESS

## THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

**Francesca Ghilotti**

*Principal supervisor:*

Associate Professor Ylva Trolle Lagerros  
Karolinska Institutet  
Department of Medicine, Solna  
Clinical Epidemiology Unit

*Opponent:*

Professor Nicholas P. Jewell  
University of California, Berkeley  
School of Public Health  
Department of Statistics

*Co-supervisors:*

Professor Torbjörn Åkerstedt  
Karolinska Institutet  
Department of Clinical Neuroscience

*Examination board:*

Paulina Nowicka  
Uppsala University  
Department of Food, Nutrition and Dietetics

Professor Rino Bellocco  
University of Milano-Bicocca  
Department of Statistics and Quantitative  
Methods  
Karolinska Institutet  
Department of Medical Epidemiology and  
Biostatistics

Christian Benedict  
Uppsala University  
Department of Neuroscience

Björn Pasternak  
Karolinska Institutet  
Department of Medicine, Solna  
Clinical Epidemiology Unit

Associate Professor Amelie Plymoth  
Karolinska Institutet  
Department of Medical Epidemiology and  
Biostatistics

Professor Anna Thorson  
Karolinska Institutet  
Department of Global Public Health



*To my family – for always believing in me*



## Abstract

The immune system plays a fundamental role in defending the human body from external pathogens. Some individuals, however, have a weaker immune system leading to more frequent infections. Infections are associated with high burden to society, both in terms of economic costs and morbidity. In this thesis we limited our attention to observational studies to investigate the association between lifestyle and infections. Our findings could be used in support of public health interventions to reduce the susceptibility to infections.

Study I examined the association between physical activity and the occurrence of upper respiratory tract infections (URTI) in a cohort of about 2,000 employed adults. Sleep duration and sleep quality were also investigated in relation to URTI. Participants filled in five questionnaires on demographics and lifestyle and were followed for 9-months for the occurrence of URTI, which was prospectively self-reported through symptoms questionnaires. Due to the excess of individuals with zero infections, hurdle regression models were used to estimate the associations under investigation. During follow-up, 1,597 URTI occurred, but our findings do not support an association of URTI with either physical activity or sleep habits.

Study II investigated the relationship between Body Mass Index (BMI) and infections requiring health care support. Participants in the study are 39,163 Swedish individuals who filled in a questionnaire in autumn 1997 and were followed through record-linkages until December 2016. Infections were identified from the Swedish national inpatient and outpatient registers. Cox proportional hazard models with age as time-scale were fitted to estimate Hazard Ratios and 95% confidence intervals. Extensions of the traditional Cox model, taking into account several infections for each individual, were used in a secondary analysis. Obesity emerged to be a risk factor for infections in both genders. In particular, obese women were at higher risk of skin infections, gastrointestinal tract infections, urinary tract infections and sepsis. Men with obesity showed an increased risk of skin infections. Findings from the extended Cox model were comparable to those from the traditional Cox model.

Study III describes the association between work-related stress and infections. Around 25,000 employed Swedish adults were followed prospectively from September 1997—when they completed a lifestyle questionnaire—until their retirement, emigration, death or December 2016, whichever occurred first. A Swedish version of the Demand-Control Questionnaire was used to assess job stress, whereas infections were identified as in Study II from the National Patient Register. In the main analysis we fitted Cox models accounting for repeated events. We found that higher job demands are associated with increased incidence of infections, in particular upper respiratory tract infections and urinary tract infections. On the other hand, our findings do not support the hypothesis that high job control is associated with a lower occurrence of infections. When combining demand and control dimensions into job strain, we found that workers with active jobs had an increased risk of infections compared to workers with low strain jobs. No difference was observed in workers with high strain jobs compared to those with low strain jobs.

Study IV aimed to explore the relationship between sleep characteristics and inflammatory markers, namely C-Reactive Protein (CRP), Interleukin-6 (IL-6) and Tumor Necrosis Factor  $\alpha$  (TNF $\alpha$ ) in a random sample of 319 non-pregnant women from Uppsala, Sweden. Participants underwent overnight in-home polysomnography (PSG), answered a sleep questionnaire and had blood samples collected the morning after PSG. We first used principal component analysis to reduce the dimensionality of the data and then estimated linear regression (after log-transformation of the outcomes) and quantile regression models to infer on the associations of interest. We found increased CRP levels in women presenting insomnia symptoms (difficulties maintaining sleep or early morning

awakenings), whereas sleep duration did not appear to be related with inflammation. From PSG measurements, a reduced REM sleep was associated with higher CRP levels. No association was found with the other markers of inflammation.

Taken together, results from these studies suggest that sleep, stress and obesity might influence the susceptibility to infections, whereas the role of total physical activity is less clear. With these studies we contributed to fill some of the knowledge gaps about the association between lifestyle and infections, but further studies are warranted to overcome the limitations encountered in our research. Furthermore, our studies serve as examples of the wide possibilities offered by statistical tools in the analysis of epidemiological data.



# List of publications

- I. Ghilotti, F., Pesonen, A. S., Raposo, S. E., Winell, H., Nyrén, O., Trolle Lagerros, Y., Plymoth, A. **Physical activity, sleep and risk of respiratory infections: A Swedish cohort study**  
*PloS ONE*, 2018, 13(1), e0190270.
- II. Ghilotti, F., Bellocco, R., Ye, W., Adami, H. O., Trolle Lagerros, Y. **Obesity and risk of infections: results from men and women in the Swedish National March Cohort**  
*International Journal of Epidemiology*, 2019, 48(6), 1783-1794.
- III. Ghilotti, F., Åkerstedt, T., Bellocco, R., Adami, H. O., Trolle Lagerros, Y. **Prospective study of job stress and risk of infections in Swedish adults**  
*Manuscript Submitted*.
- IV. Ghilotti, F., Bellocco, R., Trolle Lagerros, Y., Thorson, A., Theorell-Haglöw, J., Åkerstedt, T., Lindberg, E. **Relationship between sleep characteristics and markers of inflammation in Swedish women from the general population**  
*Manuscript Submitted*.

The articles will be referred to in the text by their Roman numerals, and are reproduced in full at the end of the thesis.

## Related publications

- Trolle Lagerros, Y., Hantikainen, E., Mariosa, D., Ye, W., Adami, H. O., Grotta, A., Ghilotti, F., Bellocco, R. **Cohort profile: the Swedish National March Cohort**  
*International Journal of Epidemiology*, 2016, 46(3), 795-795e.
- Åkerstedt, T., Ghilotti, F., Grotta, A., Bellavia, A., Trolle Lagerros, Y., Bellocco, R. **Sleep duration, mortality and the influence of age**  
*European Journal of Epidemiology*, 2017, 32(10), 881-891.
- Åkerstedt, T., Ghilotti, F., Grotta, A., Zhao, H., Adami, H. O., Trolle Lagerros, Y., Bellocco, R. **Sleep duration and mortality – Does weekend sleep matter?**  
*Journal of Sleep Research*, 2019, 28(1), e12712.
- Ghilotti, F., Julander, A., Gustavsson, P., Linde, A., Nyrén, O., Plymoth, A. **Cohort profile: Studies of Work Environment and Disease Epidemiology-Infections (SWEDE-I), a prospective cohort on employed adults in Sweden**  
*PloS ONE*, 2019, 14(5), e0217012.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background</b>	<b>2</b>
2.1	Infections . . . . .	2
2.2	Lifestyle . . . . .	3
2.2.1	Obesity . . . . .	3
2.2.2	Physical Activity . . . . .	5
2.2.3	Sleep . . . . .	7
2.2.4	Stress . . . . .	10
<b>3</b>	<b>Hypotheses and Aims</b>	<b>12</b>
3.1	Hypotheses . . . . .	12
3.2	Aims of the thesis . . . . .	13
<b>4</b>	<b>Materials</b>	<b>15</b>
4.1	The Swedish Cohorts . . . . .	15
4.1.1	Studies of Work Environment and Disease Epidemiology-Infections (SWEDE-I) . . . . .	15
4.1.2	The Swedish National March Cohort (SNMC) . . . . .	16
4.1.3	Sleep and Health in Women study (SHE) . . . . .	17
4.2	The Swedish National Registers . . . . .	18
4.2.1	The Swedish Cancer Register . . . . .	18
4.2.2	The Cause of Death Register . . . . .	18
4.2.3	The Total Population Register . . . . .	19
4.2.4	The National Patient Register . . . . .	19
4.2.5	The longitudinal integrated database for health insurance and labour market studies (LISA) . . . . .	20
<b>5</b>	<b>Methods</b>	<b>21</b>
5.1	Statistical Methods . . . . .	21
5.1.1	The Linear Regression Model . . . . .	21
5.1.2	The Quantile Regression Model . . . . .	22
5.1.3	Regression Models for Count Data . . . . .	23
5.1.4	The Cox Proportional Hazard Model . . . . .	25

5.1.5	Recurrent Event Survival Analysis . . . . .	27
5.1.6	Principal Component Analysis . . . . .	29
5.2	Overview of Methods by Study . . . . .	30
5.2.1	Study I . . . . .	30
5.2.2	Study II . . . . .	32
5.2.3	Study III . . . . .	34
5.2.4	Study IV . . . . .	36
<b>6</b>	<b>Main Results</b>	<b>38</b>
6.1	Habitual sleep, Physical activity and URTI (Study I) . . . . .	38
6.2	Obesity and Infections (Study II) . . . . .	39
6.3	Job Stress and Infections (Study III) . . . . .	42
6.4	Sleep and Inflammation (Study IV) . . . . .	44
<b>7</b>	<b>Discussion</b>	<b>47</b>
7.1	Interpretation of Findings . . . . .	47
7.1.1	Obesity . . . . .	47
7.1.2	Physical Activity . . . . .	48
7.1.3	Sleep . . . . .	49
7.1.4	Stress . . . . .	51
7.2	Methodological Considerations . . . . .	52
7.2.1	Validity . . . . .	52
7.2.2	Precision . . . . .	58
7.3	Ethical Considerations . . . . .	58
<b>8</b>	<b>Conclusions</b>	<b>60</b>
8.1	Obesity and Infections . . . . .	60
8.2	Physical Activity and Infections . . . . .	60
8.3	Sleep and Infections . . . . .	61
8.4	Stress and Infections . . . . .	61
<b>9</b>	<b>Future Research</b>	<b>62</b>
<b>10</b>	<b>Acknowledgements</b>	<b>63</b>
	<b>References</b>	<b>67</b>

# List of abbreviations

AG	Andersen-Gill
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CRP	C-Reactive Protein
DAG	Directed Acyclic Graph
DCQ	Demand-Control Questionnaire
EEG	Electroencephalography
HPA	Hypothalamus-pituitary-adrenal
HR	Hazard Ratio
ICD	International Classification of Disease
IL	Interleukin
IRR	Incidence Rate Ratio
LISA	Longitudinal integrated database for health insurance and labour market studies
MET	metabolic energy turnover
NPR	National Patient Register
NREM	Non-Rapid Eye Movement
OLS	Ordinary Least Squares
PA	Physical Activity
PCA	Principal Component Analysis
PH	Proportional Hazards
PIN	Personal Identification Number
PSG	Polysomnography
PWP	Prentice, William, Peterson
REM	Rapid Eye Movement
SNMC	Swedish National March Cohort
SNS	Sympathetic Nervous System
SHE	Sleep and Health in Women study
SSYK	Swedish Standard Classification of Occupations
SWEDE-I	Studies of Work Environment and Disease Epidemiology-Infections
SWS	Slow Wave Sleep
TNF	Tumor Necrosis Factor

TPR	Total Population Register
URTI	Upper Respiratory Tract Infection
WC	Waist Circumference
WHO	World Health Organization
WHR	Waist-to-Hip Ratio

# Chapter 1

## Introduction

The immune system plays a fundamental role in human lives. Working closely with other systems—including the endocrine system, the nervous system, and the muscular system—it defends human body from external pathogens.

Underlying medical conditions, such as cancer, might negatively influence the ability of the immune system to protect against infections. For reasons still to discover, some healthy people have a weaker immune system than their healthy pairs causing the former to be more susceptible to the seasonal flu for example, or more prone to frequent infections.

It has been hypothesized that lifestyle factors come into play here. Obesity, physical activity, sleep and stress are potential modulators of immune functions and a dysregulated immune system is one of the conditions leading to higher occurrence of infections. Obesity is thought to modulate the immune response through adipokines secreted by the adipose tissue (Falagas and Kompoti, 2006); exercise might boost the immune system through increased production of anti-inflammatory cytokines, NK-cells and neutrophils (Nieman and Wentz, 2019); effects of sleep on immune responses are mediated by the sympathetic nervous system and the neuroendocrine system (Besedovsky et al., 2012); finally long-term stress, through the release of cortisol hormone, is also believed to be related to the immune system functioning (Glaser and Kiecolt-Glaser, 2005).

Studies published so far have mostly focused on small or restricted groups of individuals. Studies on sleep are mainly conducted in experimental settings and focused principally on severe sleep deprivation; studies on physical activity have mostly investigated exercise or leisure time physical activity rather than total physical activity (i.e. including sedentary behaviors); studies on stress are chiefly cross-sectional and rely on self-reported outcomes. Moreover, these lifestyle factors have mostly been examined in relation of respiratory infections. Other community-acquired infections, despite less frequent than respiratory ones, are also related to high costs to society and high burden in terms of morbidity (Christensen et al., 2009), therefore the whole picture is warranted.

As there is still uncertainty, with this thesis we aimed to investigate more on the role of lifestyle factors on infection susceptibility using information from everyday life. To accomplish the aim, four epidemiological studies were conducted.

# Chapter 2

## Background

### 2.1 Infections

Infectious diseases are caused by microorganisms that invade the body's tissues. The main characteristic of these diseases is their ability to spread from one person to another. Pathogenic microorganisms responsible for infections are for example bacteria and viruses. Bacteria are tiny single-cell microorganisms that can live in any kind of environment and survive on their own, whereas viruses cannot survive without a host and can only reproduce by attaching themselves to cells. Though the symptoms deriving from one or the other infection type could be similar and include fever, coughing, running nose, fatigue and vomiting, the treatments are quite different. Bacterial infections can be treated with antibiotics, while treatments for viral infections are mostly aimed to improve symptoms. Some viral infections, such as HIV, hepatitis, herpes viruses, and influenza A and B can be treated with antiviral drugs which interfere with the production of viral DNA inhibiting virus replication (Razonable, 2011). Alternatively, viruses can also be prevented through vaccination.

The immune system protects our body from these pathogens through two main components: the innate and the adaptive immune system (Parkin and Cohen, 2001). The first one is the first line of defense and it activates a general immune response, such as inflammation; the second one is responsible for a highly specific response to a particular pathogen and for immunological memory. Cytokines, small proteins secreted by immune systems cells, are important mediators of the immune response. Among the pro-inflammatory cytokines, namely cytokines that enhance the inflammatory response, are interleukin 1 (IL-1), IL-6 and Tumor Necrosis Factor  $\alpha$  (TNF $\alpha$ ). When an inflammatory reaction begins, the level of the C-reactive protein (CRP) rise. As a result, CRP is used as marker for inflammation in the body. Given the difficulties in early identification of infection's nature, CRP levels are also used to distinguish among viral and bacterial infections. CRP usually reaches higher levels for bacterial infections, compared to infections caused by viruses (Sasaki et al., 2002).

Infections are associated with considerable morbidity and mortality. The most recent case is that of the novel coronavirus (Covid-19) outbreak, which has recently been recog-



nized as a pandemic from the World Health Organization (WHO). Within a few months from its identification (first case was identified in Wuhan, China on December 31, 2019), it has already caused more than 13,000 deaths worldwide. Covid-19 has spread in 171 countries so far, affecting over 300,000 people (data updated on March 21, 2020).

Lower respiratory tract infections, responsible for 2.4 million deaths worldwide in 2016, were a leading infectious cause of mortality (GBD 2016 Lower Respiratory Infections Collaborators et al., 2018). Hospitalizations for infectious diseases have increased over time with subsequent high medical costs (Christensen et al., 2009). Respiratory tract infections is the main reason to visit healthcare in the Nordic countries (Grimsmo et al., 2001). Adults experience an average of 2-3 episodes of common cold each year with subsequent high cost to society due to absenteeism from work and productivity loss (Hellgren et al., 2010).

Due to the consistent burden caused by infections in terms of morbidity, mortality, and costs to society, there is an emerging need to find modifiable lifestyle factors that could reduce infection risk.

## 2.2 Lifestyle

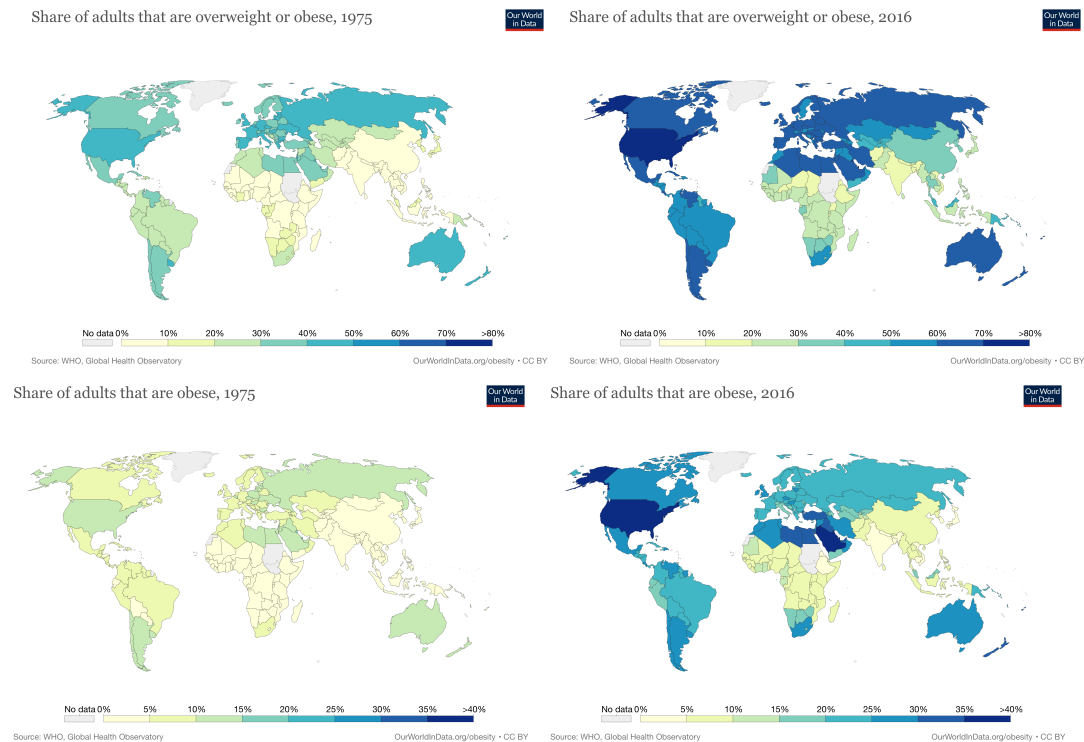
### 2.2.1 Obesity

Obesity is a medical condition characterized by excessive fat accumulation and it is associated with a higher susceptibility to comorbidities and an increased mortality. The cause of obesity is attributable to an energy imbalance between energy consumed and energy expended (Hill et al., 2012), namely diminished physical activity levels and dietary changes towards energy-dense diets. For many years obesity has been considered a lifestyle choice; however, starting in 1998 the National Institutes of Health declared obesity a disease (NIH, 1998), followed by the Obesity Society in 2008 (The Obesity Society, 2008; Jastreboff et al., 2019).

Obesity is commonly measured using the body mass index (BMI), calculated as weight in kilograms divided by height in meters squared. According to the standard classification from the WHO (WHO, 2000), individuals are classified into underweight ( $\text{BMI} < 18.5 \text{ kg/m}^2$ ), normal weight ( $\text{BMI} \geq 18.5$  to  $< 25 \text{ kg/m}^2$ ), overweight ( $\text{BMI} \geq 25$  to  $< 30 \text{ kg/m}^2$ ), obesity ( $\text{BMI} \geq 30$  to  $< 35 \text{ kg/m}^2$ ), severe obesity ( $\text{BMI} \geq 35$  to  $< 40 \text{ kg/m}^2$ ) and morbid obesity ( $\text{BMI} \geq 40 \text{ kg/m}^2$ ). Despite the fact that BMI is an easy-to-calculate measure and equally applied in males and females, it also has some drawbacks. Because it does not distinguish between lean mass and fatty tissue, the same BMI value may correspond to two persons with very different body compositions. For this reason, other measures such as waist circumference (WC), or waist-to-hip ratio (WHR) have been developed to more accurately measure abdominal fat (WHO, 2000).

Obesity is one of the largest health problem in the world. Not only is obesity an issue in rich countries, but it has also become a major public health concern in middle-income and developing countries (see Figure 2.1). Overall, in 2016 the WHO estimated that 11% of

men and 15% of women in the world were obese (WHO, 2018b), with its prevalence nearly tripled in the last four decades (NCD Risk Factor Collaboration et al., 2016). Obesity is included among the global non-communicable disease targets for 2025 by the WHO (WHO, 2013). However, if this increasing trend continues, there is no chance to meet the WHO goal of halting the rise in obesity by 2025. In 2017, high BMI was the fourth risk factor for mortality, accounting for 4.7 million deaths worldwide; in Sweden, 8.7 thousand deaths were attributable to obesity in 2017 (GBD 2017 Risk Factor Collaborators et al., 2018).



**Figure 2.1:** Proportion of overweight and obese individuals worldwide: The maps show the share of adults with  $BMI \geq 25$  or  $BMI \geq 30$  across countries in 1975 and 2016. Reproduced under the Creative Commons BY license from <https://ourworldindata.org/obesity> (Ritchie and Roser, 2020).

Obesity induces chronic low-grade inflammation (Lee et al., 2013) and it has been associated with an increased risk and worse outcomes of infections (Huttunen and Syrj nen, 2013). Several mechanisms are involved in the underlying pathway between this medical condition and the increased susceptibility to infections.

First, adipose tissue secretes leptin and adiponectin, adipokines which modulate the immune response (Falagas and Kompoti, 2006; Karlsson and Beck, 2010). Leptin, whose levels are increased in obese subjects, is a pro-inflammatory factor, while adiponectin, whose levels are decreased in obese subjects, is an anti-inflammatory factor. Leptin resistance induced by hyperleptinaemia may explain the impaired immunity encountered in individuals with obesity (Milner and Beck, 2012). Second, comorbidities related to the presence of obesity, such as diabetes type 2, might contribute to the immune response dysregulation (Huttunen and Syrj nen, 2013). Third, altered pharmacokinetic and pharmacodynamics parameters are encountered in obese subjects, which leads to difficulties in the appropriate

dosing of antimicrobials and worse outcomes of infections (Huttunen and Syrjänen, 2013). Lastly, respiratory dysfunction, diminished peripheral blood perfusion and skin folds are factors related to obesity which may induce an increased risk for respiratory tract infections, abscesses and skin infections, respectively (Kaspersen et al., 2015).

Some reviews tried to put together the up-to-date findings in regards of the relationship between obesity and infections (Huttunen and Syrjänen, 2013; Dobner and Kaser, 2018). Associations with skin infections are well established, but results on some other specific infection types are less clear. Urinary tract infections have been found to be associated with obesity to different extents in males and females; inconsistent results were found for sepsis and respiratory tract infections (often studied without distinguishing between infections of the lower or upper respiratory tract); limited data are available for gastrointestinal tract infections. As the relationship of obesity with many infectious diseases is still under discussion, further studies are warranted.

### **2.2.2 Physical Activity**

Physical activity is defined as movements of human body that require energy expenditure. It should not be confused with exercise; the latter refers to a planned activity aimed to improve or maintain physical fitness (Caspersen et al., 1985). Physical activity is a key modifiable lifestyle factor in preventing many non-communicable diseases (WHO, 2010). Regular physical activity is also associated with better quality of life, mental health and well-being (Scully et al., 1998).

Recommendations on the minimum level of physical activity have been released by the WHO. Adults aged 18-64 years should do at least 150 minutes each week of moderate-intensity aerobic activity, or 75 minutes each week of vigorous-intensity activity, or perform a combination of both forms of physical activity (WHO, 2010). In 2016, 32% of men and 42% of women in high income countries did not reach the minimum level of physical activity recommended by WHO and were therefore classified as physically inactive (Guthold et al., 2018). The prevalence was lower in low income countries (13% of men and 19% of women). These estimates, however, derive from self-reported data and might underestimate the dimension of the problem.

In 2017, the lack of physical activity was deemed responsible for 1.3 million deaths worldwide (GBD 2017 Risk Factor Collaborators et al., 2018). This fact might be explained by the direct link existing between physical inactivity and the increased risk of the world's major non-communicable diseases, such as cardiovascular diseases, cancer, and diabetes (Lee et al., 2012; Knight, 2012). Due to the worldwide urgency of the problem, WHO launched a Global Action Plan to reduce 15% the global prevalence of physical inactivity by 2030, keeping 2016 as baseline (WHO, 2018a).

Beside physical inactivity, another emerging problem is that of a sedentary behaviour. Sedentary lifestyle has become more widespread as a consequence of transformations in the workplace (e.g. smart-working), innovations in transportation (e.g. elevators, escalators,

e-bikes, e-scooters) and domestic-entertainment technologies (e.g. streaming platforms, remote controls) (Owen et al., 2010). Even if one is physical active according to WHO guidelines, it is important to consider how the rest of waking hours are spent, because effects of prolonged sitting have been found independently of physical activity level (Bauman et al., 2011).

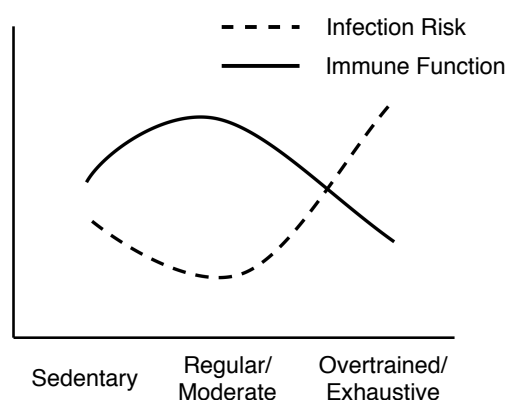
In 1993, Ainsworth published a comprehensive list of physical activities with their intensities expressed as energy cost (Ainsworth et al., 1993). Based on previous publications, each activity was coded with a specific value of metabolic energy turnover (MET). One MET is defined as the energy expenditure of 1 kcal/kg body weight per hour. It is considered the reference metabolic rate and it corresponds to quiet sitting; all other activities are expressed as multiples of one MET. Light-intensity physical activity equals <3.0 METs, moderate-intensity between 3.0 and 6.0 METs and vigorous-intensity >6.0 METs (Table 2.1) (WHO, 2014b). Since then, different self-reported questionnaires for physical activity levels have been proposed with the purpose of deriving the total energy expenditure during an ordinary 24-hours day.

**Table 2.1:** General physical activities by level of intensity

<b>Moderate Activity 3.0 to 6.0 METs</b>	<b>Vigorous Activity Greater than 6.0 METs</b>
Brisk walking	Hiking
Dancing	Running or Jogging
Yoga	Bicycling fast
Gardening	Aerobics
Housework and domestic chores	Competitive sports and games
Actively playing with children	(e.g. Football, Soccer, Volleyball, Basketball)
Swimming (recreational)	Fast swimming
Hunting	Carrying heavy loads
	Heavy shovelling or digging ditches

Epidemiological evidence suggests a link between exercise and the immune system (Romeo et al., 2010; Nieman and Wentz, 2019). In 1994 Nieman described the relationship between exercise and upper respiratory tract infections (URTI) to be J-shaped (Nieman, 1994). The model suggests that the risk of URTI decreases when engaging in moderate-intense physical activities compared to a sedentary behavior, but during high-intense exercise the risk can be even higher than being sedentary. A few years later the “Inverted J-Hypothesis” was proposed. Namely, it was hypothesized that a moderate training might enhance the immune response, whereas a prolonged high intensity exercise might lead to immunosuppression (Woods et al., 1999) (see Figure 2.2). As a result, the effects of exercise on the immune system depend on several factors among which the intensity, duration and type of exercise.

The higher risk of infections observed after strenuous exercise could be explained by the “open window” theory (Nieman, 1994). The “open window” is considered the period of time between 3 and 72 hours immediately following a high intensity prolonged training; during



**Figure 2.2:** Inverted J-Hypothesis, adapted from Wood (Woods et al., 1999)

this span of time individuals experience an altered immunity. This state is characterized by an increased number of leukocytes immediately after exercise (Nieman, 1994), likely due to epinephrine and cortisol changes. This state is only transient and after a few hours the immune function is back to pre-exercise levels.

Although the relationship of exercise on the immune response is well established, the enduring effect of a regular moderate physical activity on our immune system is still under investigation (Romeo et al., 2010; Nieman and Wentz, 2019). It has been argued that the immune function in non-athletes and athletes during resting state is not dissimilar, supporting evidence that the changes on the immune system are only transitory (Nieman and Pedersen, 1999). Despite that, cross-sectional studies showed a long-term effect of regular physical activity on serum CRP levels, which were constantly lower in more physically active individuals (Kasapis and Thompson, 2005). Physical activity might act both directly by altering the production of cytokines in muscle, fat and mononuclear cells and indirectly by enhancing oxygenation, improving endothelial function, reducing body weight and increasing insulin sensitivity (Kasapis and Thompson, 2005; Nieman and Wentz, 2019).

The published studies investigating the effect of physical activity on infections have mainly focused on respiratory infections, in particular of the upper respiratory tract. Despite that some observational studies showed an inverse association between higher physical activities levels and respiratory infections (Matthews et al., 2002; Fondell et al., 2011; Nieman et al., 2011), a large Finnish study did not report any association (Hemilä et al., 2003), and no increased risk was found in a recent Cochrane systematic review (Grande et al., 2015). The role of regular physical activities on infection occurrence is not yet established and further studies are needed.

### 2.2.3 Sleep

Sleep is an important and potentially modifiable factor that may influence human health. The awareness of its role as an essential component of health has globally increased due to the rising number of publications relating both short and long sleep to higher mortality

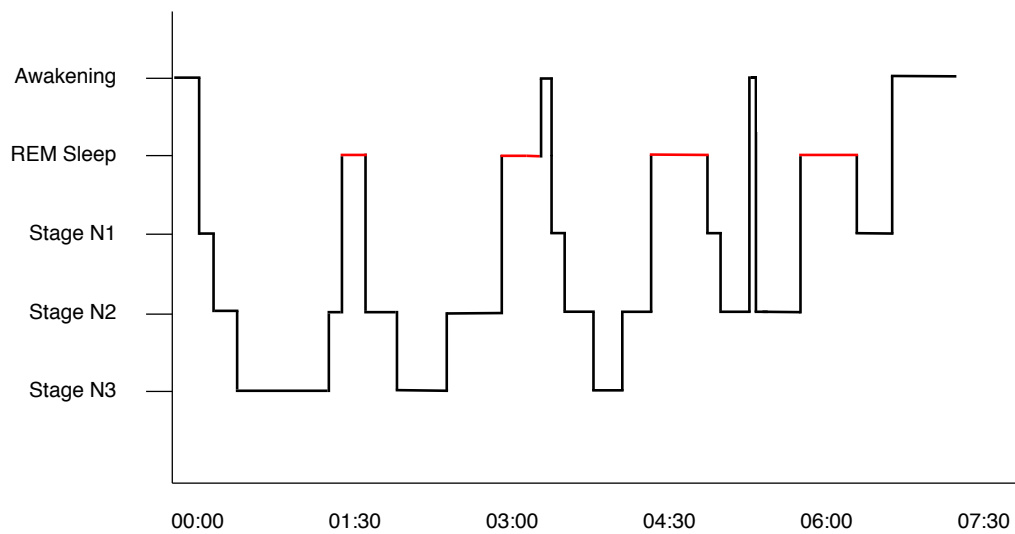
rates and increased risk of several diseases (Ferrie et al., 2011; Itani et al., 2017; Jike et al., 2018). The vital importance of sleep becomes obvious if one considers that humans spend around one third of their lives sleeping.

Sleep is a reversible behavioral state characterized by reduced consciousness and decreased responsiveness to environmental stimuli (Shneerson, 2009). It is sometimes defined as the opposite of wakefulness, though the margin between these two stages is seldom well defined. The sleep-wake cycle, described for the first time in the 80's (Borbély, 1982), is regulated by two body systems: the circadian system and the sleep-wake homeostasis, wherein the homeostatic sleep pressure increases during wakefulness and declines during sleep. Circadian rhythms, as suggested by the Latin origin of the word (*circa* meaning “around” and *diem* meaning “day”), are physiological processes with a 24-hour cycle. Circadian rhythms are regulated by an endogenous biological clock situated in the Suprachiasmatic Nucleus within the hypothalamus. This group of cells in the brain is sensitive to exogenous stimuli, the so-called Zeitgebers (from German terms *zeit* meaning “time” and *geber* meaning “giver”—literally “time givers”) such as light, temperature and noise, which enables the biological clock to be entrained or reset. The greatest example of its flexibility is the adaption of the body's internal circadian clock to new time zones when travelling.

The invention of electroencephalography (EEG) in 1924 paved the way to the study of the activity of human brain during sleep. Nowadays, it is well known that normal sleep architecture is characterized by two main subsequent phases: Non-Rapid Eye Movement (NREM), and Rapid Eye Movement (REM), each of which is characterized by its own typical features. NREM sleep is further subdivided into three stages (N1, N2, and N3). The first two are light sleep; the latter, also known as slow wave sleep (SWS) due to the presence of delta waves in EEG, consists of deep sleep. While in NREM sleep, the body temperature, blood pressure, and heart rate decrease and our body spend less energy than when awake. During REM sleep, these parameters are irregular, muscles are paralyzed and individuals are likely to dream vividly with as much as energy expenditure reported as during wakefulness (Porkka-Heiskanen et al., 2013). The alternation of these four sleep stages is commonly referred to as sleep cycles. A sleep cycle lasts for about 90 minutes and a total of four to six cycles occur each night (Shneerson, 2009) (see Figure 2.3).

The sleep and the immune system are reciprocally linked (Bryant et al., 2004; Besedovsky et al., 2019; Irwin, 2019). Sleep deprivation impairs the immune response and increases the susceptibility to infections; conversely, the response to an infection might alter sleep, inducing increased somnolence. Partly responsible for this latter mechanism are the pro-inflammatory cytokines (in particular TNF and IL-1) produced during the acute-phase response to infections. TNF and IL-1 are substances involved in the homeostatic regulation of sleep and enhance sleep.

Sleep has in itself a role on the immune system and this direction of the association was of interest in our studies. While sleeping, peaks of the pro-inflammatory cytokines and suppression of anti-inflammatory hormones have been observed, which together lead to a boosted immune activation (Besedovsky et al., 2012). The effect of sleep on the im-



**Figure 2.3:** A hypnogram showing sleep architecture in a normal adult

immune system is thought to be partly mediated by the neuroendocrine system (Bryant et al., 2004). Both the neuroendocrine system and some immune system components have a circadian rhythmicity, supporting the evidence of a link between these systems. Sleep increases the secretion of anabolic hormones which regulate the immune function such as growth hormone and prolactin, and reduce the release of catabolic hormone such as cortisol. Melatonin, whose secretion is inhibited when exposed to light, is also involved in the immune system regulation. Sleep disturbances result in repeated activations of the hypothalamus-pituitary-adrenal (HPA) axis which in turn releases cortisol that suppresses pro-inflammatory and antiviral immune responses. Another mechanism linking sleep and immunity is through the sympathetic nervous system (SNS) (Besedovsky et al., 2012; Irwin, 2015). During sleep, the SNS activity decreases; when disturbances in sleep occur, the SNS is continuously activated. It has been shown that activation of SNS alters the production of innate immune cells and leads to increased inflammation (Irwin and Cole, 2011).

Although experimental studies in animals have proven causality of the link between sleep and immune system (Besedovsky et al., 2019), a limited number of studies in humans have investigated the relationship between sleep and infection risk. In addition, studies conducted at a population level are still scarce. Two studies with experimental viral exposure showed a higher incidence of common cold in individuals who reported shorter sleep duration (Cohen et al., 2009; Prather et al., 2015a). As of observational studies, a higher risk of pneumonia (Patel et al., 2012) and a higher likelihood of reporting infection occurrence in the previous month (Prather and Leung, 2016) were found in subjects with short sleep. The relationship of long sleep duration on infections, however, is less clear and studies are needed (Besedovsky et al., 2019). With regards to sleep and inflammatory markers, a greater number of studies has been conducted, but most of them rely on self-reported sleep, and those with objective measurements are usually based on a small sample size.

In addition, inconsistent results were found in respect of different inflammatory markers (Irwin et al., 2016).

With the advent of an “always-on” society, the frequent night-time use of screens, and the increased availability of 24/7 services which expose more subjects to shift-work, the quality and the duration of our sleep might be compromised. According to the Centers for Disease Control and Prevention (CDC), sleep disorders constitute a public health epidemic. In 2014, the age-adjusted prevalence of sleep duration <7 hours was 35% in the US (CDC, 2014). In Sweden, the prevalence of women complaining about sleep problems has almost doubled over three decades, going from 18% and 22% in 1968, to 32% and 42% in 2004, among 38- and 50-years old, respectively (Rowshan Ravan et al., 2010).

Given the worldwide increase of sleep problems, it is of importance to investigate the relationship between sleep and immune system by filling the knowledge gaps identified.

#### 2.2.4 Stress

Stress was defined for the first time in 1936 by Hans Selye who described it as “the non-specific response of the body to any demand for change” (Tan and Yip, 2018). The external stimulus that causes stress to an organism is called stressor. Both psychological and physiological changes are activated in response to a stressor.

Stressors have been classified in five categories on the basis of two dimensions: duration and course (e.g. occasional or uninterrupted) (Elliott, 1982). *Acute time-limited stressors* involve challenges such as mental calculations or public speaking. *Brief naturalistic stressors* involve real-life short-term challenges such as academic examinations. *Stressful event sequences* arise in connection to an important event, such as the death of a close person or loss of job, which originate a chain of related challenges. The individuals, however, intuitively know that at some point in the future these challenges will wane. *Chronic stressors* pervade enduringly a person’s life. As a result, the affected person does not know whether or when the challenge will end. Examples of this type of stressors include the loss of a limb or providing care for a spouse with a neurodegenerative disease. *Distant stressors* refer to traumatic events that occurred in the distant past, but still have cognitive and emotional consequences in the present, such as having been raped or tortured as a child.

Individuals exposed to high levels of psychological stress are more likely to have worse health conditions than non-stressed persons; in particular, high stress is known to be related to cardiovascular diseases and mental disorders (Kivimäki and Steptoe, 2018). There is also increasing evidence that stress makes individuals more susceptible to infections (Glaser and Kiecolt-Glaser, 2005).

The effect of stress on the immune system is dependent on stressor type. According to a biphasic model, the immune response is enhanced by acute stress, and it is suppressed by chronic stress (Dhabhar and McEwen, 1997). Findings in support of this model were reported in a meta-analysis of more than 300 studies (Segerstrom and Miller, 2004). The two main mechanisms through which stress seems to affect the immune system are through



nerve fibers and hormonal responses (Cohen, 1996; Reed and Raison, 2016). The sympathetic nervous system, activated in response of a stressor, releases into lymphoid organs norepinephrine and epinephrine, which in turn promote immune functions (Padgett and Glaser, 2003). Psychological stress can activate the hypothalamus-pituitary-adrenal axis and therefore cause the release of cortisol with adverse immunological changes (Glaser and Kiecolt-Glaser, 2005). Additionally, behavioral changes such as smoking, alcohol use, not eating or sleeping properly might be mediators in the pathway linking stress to the immune system (Cohen, 1996; Segerstrom and Miller, 2004).

According to the sixth European working conditions survey conducted by Eurofound (the European foundation for the improvement of living and working conditions) in 35 European countries, one-third of workers work to tight deadlines and at high speed (Eurofound, 2017). Work intensity is one of the main factor promoting job-related stress. Not only stress in the workplace influence the well-being of individuals, but it is also associated with a considerable economic burden to society. The high costs are both direct (caused by medical costs) and indirect (due to production loss) and were estimated to be € 26.47 billion within the EU-15<sup>1</sup> for 2014 (EU-OSHA, 2014; Hassard et al., 2018).

Despite the dimension of the problem, few studies have investigated the relationship between work stressors and risk of infections. Additionally, most of the existing studies are cross-sectional, and measured self-reported outcomes or surrogate end-points (e.g. plasma fibrinogen concentration, reduced NK-cell activity) (Morikawa et al., 2005; Boscolo et al., 2009, 2012). Despite the link found between job stress and common cold (Mohren et al., 2001), further studies are needed to examine the relationship with other infections.

---

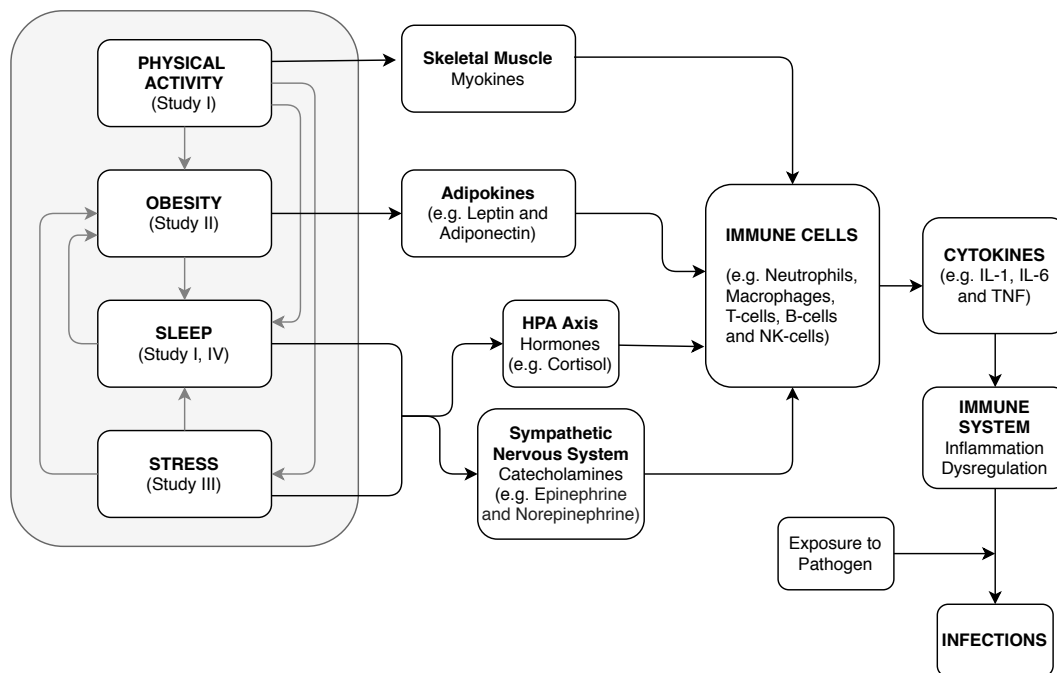
<sup>1</sup>Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom.

# Chapter 3

## Hypotheses and Aims

### 3.1 Hypotheses

The immune system is interconnected with other systems in the body—e.g. the endocrine system, the nervous system and the muscular system—as briefly described in the previous section. We hypothesized common mechanisms underlying the association between lifestyle and infections, irrespectively of the severity of the outcome. Therefore, the outcomes presented in the four studies included in the thesis differ. Common cold and severe infections, are both resulting from a dysregulated immune system in which lifestyle factors are one of the possible modulators of the immune cells production (see Figure 3.1).



**Figure 3.1:** Illustration of some possible mechanisms for an association between lifestyle and infections.

Figure 3.1 illustrates some of the possible underlying mechanisms linking lifestyle and infections. It is not our intention to give an exhaustive picture of all possible arrows existing between lifestyle and infections or between lifestyle factors themselves. It merely illustrates the fact that different lifestyle habits are connected to each other. For example, regular physical activity leads to a reduction of adipose tissue, it might improve sleep and decrease stress levels. The link between obesity and sleep might be thought as bidirectional. Obesity increases the chance of obstructive sleep apneas, whereas lack of sleep might impact appetite regulation. Individuals exposed to high stress levels are likely to be sleep deprived and more likely to gain weight, as they consume more food high in fat and sugars to relieve stress.

Concerning specific hypotheses, we assumed that a low level of physical activity has a negative effect on the immune system. There are previous studies, but most of them focused on exercise instead of total physical activity. The relationship between obesity and infections is quite well established, but studies distinguishing among infection types are scarce. We hypothesized that individuals with obesity are more susceptible to infections, and their susceptibility might vary across infection subgroups. As for stress, few studies are available and most of them focused on respiratory infections. We assumed that jobs with high demands would lead to an increased incidence of infections, and this effect might be mitigated by a high control. Finally, the link between sleep and respiratory infections has mostly been studied in experimental studies. Based on results from these studies we hypothesized that short sleep is associated with a higher risk of infections, and that the effect could be modified by sleep quality. Also, we assumed that worse sleep negatively affects the immune system.

## 3.2 Aims of the thesis

The overall aim of this doctoral thesis is to extend previous knowledge about the lifestyle factors that are relevant for the occurrence of infections, viral as well as bacterial, with particular focus on the role of obesity, physical activity, sleep and stress. This work is also intended to be a source of inspiration for future researchers giving practical examples of the use of more sophisticated statistical tools than commonly used.

More specifically, the aims were:

- To investigate whether physical activity, sleep duration and sleep quality are associated with the incidence of upper respiratory tract infections (Study I)
- To assess and quantify the association between obesity and risk of a wide range of community acquired infections requiring health care contact (Study II)
- To evaluate whether job stress, assessed through the demand-control questionnaire, increases the risk of community-acquired infections requiring health care contact (Study III)

- To study the relationship between sleep characteristics, both self-reported and objectively measured, and markers of inflammation (Study IV)

# Chapter 4

## Materials

The studies included in this thesis are all observational and were conducted based on data from three different Swedish cohorts. Study I is based on data from “Studies of Work Environment and Disease Epidemiology-Infections” (SWEDE-I) (Ghilotti et al., 2019); Studies II and III used data from the “Swedish National March Cohort” (SNMC) (Trolle Lagerros et al., 2017) linked to several Swedish National Registers thanks to the availability of the unique Personal Identification Number (PIN) (Ludvigsson et al., 2009); Study IV investigated a sample of women included in the “Sleep and Health in Women study” (SHE) (Theorell-Haglöw et al., 2006). In this chapter, cohorts and registers used in this thesis are described.

### 4.1 The Swedish Cohorts

#### 4.1.1 Studies of Work Environment and Disease Epidemiology-Infections (SWEDE-I)

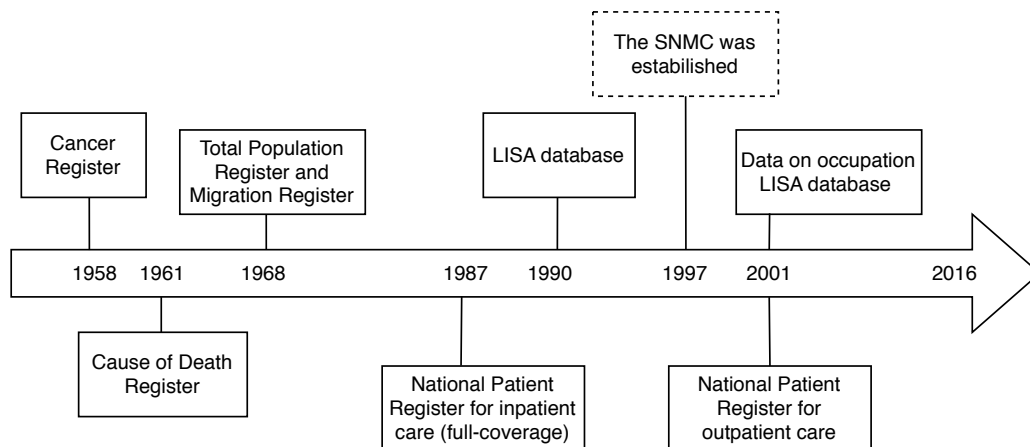
The SWEDE-I cohort was set up in August 2011 to investigate work-related factors linked to a higher occurrence of viral infections (Ghilotti et al., 2019). Invitations were sent to a random sample of 14,008 employed individuals who both worked and lived in the city Eskilstuna, in central Sweden. After excluding individuals who did not answer to the initial invitation and one reminder, who did not meet study criteria (not currently working), and who were unwilling to participate, a total of 2,237 subjects were included in the cohort. Participants filled in five detailed questionnaires on demographics, anthropometric measures, work place, work tasks, health status, physical activity and dietary habits and were asked to self-report any onset of fever, respiratory infections and gastroenteritis during 9-month follow-up. At each event, participants filled in a symptom questionnaire and submitted a nasal swab to a laboratory to be analyzed for 14 different viruses. Baseline characteristics of the SWEDE-I cohort, in comparison with the other two cohorts used in this thesis, are reported in Table 4.1.

#### 4.1.2 The Swedish National March Cohort (SNMC)

In September 1997, the Swedish Cancer Society organized a four-days national fundraising event which took place in around 3,600 cities and villages throughout the country. Beside fundraising galas and scientific programmes on television, there was also a 1-10 km walk for cancer, named the National March, with several different starting points across Sweden. Partaking costed 50 Swedish crowns which was a donation for cancer research. People coming to the event (not necessarily participants to the walk) were asked to fill in a 36-page questionnaire called "An Hour for Research" covering demographics, lifestyles and health. The questionnaire was available at each starting point of the walk and had to be returned in one of the local supermarkets. The questionnaires were then transported by milk trucks to Statistics Sweden for scanning. In total, 43,880 questionnaires were returned; after removing individuals with incomplete/inconsistent PIN, the final cohort consisted of 43,865 men and women.

Thanks to the unique PIN provided by each participant, the cohort was linked to several nationwide and well-validated registers (Cancer Register, Cause of Death Register, Total Population Register, Migration Register, National Patient Register, and Swedish Longitudinal Integrated Database for Health Insurance and Labour Market Studies database [LISA]; see Figure 4.1) and followed-up through record-linkage until December 2016.

The SNMC was set up to study the role of lifestyle factors, mainly physical activity, in the development of cancer, cardiovascular diseases and other chronic conditions. Baseline characteristics of the SNMC cohort, in comparison with the other two cohorts used in this thesis, are reported in Table 4.1.



**Figure 4.1:** Timeline of the establishment of the Swedish National Registers used in this thesis

**Table 4.1:** Comparison between SWEDE-I, SNMC and SHE in terms of cohort design, age and sex distribution, educational level, health status, body mass index and smoking status.

	<b>SWEDE-I</b>	<b>SNMC</b>	<b>SHE</b>
Setting	Employed individuals living and working in Eskilstuna, Sweden	Participants of a cancer fundraising event organized in 3,600 sites throughout Sweden	Random sample of women selected from the population registry of Uppsala, Sweden
Participants	2,237	43,865	400
Year of start	2011	1997	2002
Follow-up	9 months	19 years	–
Age, median (IQR)	46 (39-56)	52 (38-64)	52 (43-58)
Sex, % Female	59.0	64.3	100
Education, %			
≤9 years	8.9	38.4	17.7 <sup>a</sup>
11-13 years	33.2	34.4	31.5 <sup>a</sup>
≥15 years	57.9	27.2	50.8 <sup>a</sup>
Health, %			
Very good/Good	80.9	82.1	83.2
Neither good nor bad	16.5	15.1	14.2
Bad/Very bad	2.6	2.8	2.6
BMI, %			
<18.5 kg/m <sup>2</sup>	0.6	1.9	0.3
18.5–24.9 kg/m <sup>2</sup>	47.0	59.2	43.7
25–29.9 kg/m <sup>2</sup>	37.7	31.8	34.5
≥30 kg/m <sup>2</sup>	14.7	7.1	21.5
Current smokers, %	11.2	7.8	21.1

<sup>a</sup> Data on education were collected in 2010, in a third phase of the study.  
IQR; Interquartile Range. BMI; Body Mass Index.

### 4.1.3 Sleep and Health in Women study (SHE)

In the year 2000, a total of 10,000 women selected from the population register of the Swedish city of Uppsala were invited to participate in the Sleep and Health in Women study (Theorell-Haglöw et al., 2006). Information on sleep habits and health status were collected through a self-reported questionnaire, which has been returned by a total of 7,051 women (71.6%).

In a second phase, a random sample (with over-representation of habitual snorers) of 400 women was selected for home-based, full-night polysomnography (PSG) and laboratory

testing. The PSG included electroencephalogram, electrocardiogram, electrooculogram, electromyogram, airflow, respiratory effort, finger oximetry, pharyngeal sounds, and body position. The evening before the PSG women went to the sleep laboratory to pick up the PSG equipment. Before the participants could return to their homes to sleep, a research nurse applied the electrodes and gave instructions. The morning after the PSG, participants went back to the laboratory to remove the equipment and to undergo laboratory testing. Baseline characteristics of the 400 women included in Phase II of the SHE study are reported in Table 4.1.

A third phase of the study took place in 2010, when a follow-up questionnaire was sent to all the women who answered the baseline questionnaire 10 years before. In this thesis, however, only cross-sectional data from the 400 women who underwent PSG were used.

## **4.2 The Swedish National Registers**

Sweden, as well as the other Nordic countries, is well known for its long tradition of systematically collecting data in nationwide registries, as well as for its extensive use of PINs for everyday purposes. The availability of PINs enables linkages between different registries in order to conduct medical research on large cohorts with a long and virtually complete follow-up (Ludvigsson et al., 2009). The PIN, introduced in Sweden in 1947, is a 10-digit number in the format YYMMDD-XXXX consisting of the date of birth, a sex-specific three-digit birth number (odd for man and even for women) and a check-digit introduced in 1967 to verify that date of birth and the three-digit number are correct. Each individual who has resided in Sweden for at least one year has been given a PIN.

### **4.2.1 The Swedish Cancer Register**

The Swedish Cancer Register (Barlow et al., 2009), initiated in 1958, covers the whole population and it is maintained by the National Board of Health and Welfare. Every physician is obliged by law to report newly diagnosed cancer cases. The register provides detailed information on patient data, medical data (site of tumor, histological type, date of diagnosis) and follow-up data (e.g. date and cause of death, date of emigration).

The Cancer Register was used in Studies II and III to identify prevalent cancer patients at baseline, in order to obtain the Charlson's comorbidity index (Charlson et al., 1987).

### **4.2.2 The Cause of Death Register**

The Swedish Cause of Death Register contains information on date of death, the underlying cause of death, and the contributing cause(s) since 1952 (Brooke et al., 2017). The register started to be updated on annual basis from 1961, when it reached its nationwide coverage. The register is maintained by the National Board of Health and Welfare and mortality data are available for all individuals registered in Sweden at the time of death, regardless of whether they died in Sweden or abroad.



The Cause of Death Register was used in Studies II and III to retrieve dates of death of individuals included in the SNMC in order to calculate their correct follow-up time.

#### **4.2.3 The Total Population Register**

The Total Population Register (TPR) started in 1968 and contains demographic information (e.g. name, sex, date and place of birth, civil status, citizenship, date of death) on all individuals who are residing in Sweden for at least one year (Ludvigsson et al., 2016). The TPR also contains the Migration Register, through which data on emigrations and immigrations can be obtained. The TPR is maintained by Statistics Sweden and it is constantly updated by the Swedish Tax Agency, which delivers data to Statistics Sweden. The quality of the TPR is generally high, but it might decrease when the reporting of information depends on the individual (residence change or migration). When individuals fail to report emigration, over-coverage arises. This phenomenon is more frequent in residents born outside the Nordic countries.

The TPR was used in Studies II and III to retrieve dates of emigration of individuals included in the SNMC in order to censor their follow-up at the time of their first emigration outside the country.

#### **4.2.4 The National Patient Register**

The National Patient Register (NPR) consists of the Swedish Inpatient Register and the Swedish Outpatient Register. The Swedish Inpatient Register was established in 1964 when the Swedish National Board of Health and Welfare started collecting data on overnight hospitalizations in some Swedish counties; but full national coverage was not reached until 1987 (Ludvigsson et al., 2011). Since 2001, all hospital-based outpatient visits of non-admitted patients are reported in the Swedish Outpatient Register. All physicians, from private and public sectors, are obliged to report any hospital discharge and all hospital-based outpatient visits; however, primary care is yet not recorded in the NPR. The register contains information on patient-related data such as the PIN, sex, and age; information on hospital and department of admission; administrative data like admission and discharge dates; and medical data on primary and additional diagnoses at discharge. Each discharge or visit is coded according to the Swedish Revisions of the International Classification of Disease (ICD) system (ICD-7: 1964-1968; ICD-8: 1969-1986; ICD-9: 1987-1996; ICD-10: 1997 onwards). Because of the availability of the PIN, the long follow-up and the high validity for most of the diagnoses (positive predictive value ranging from 85% to 95%), the NPR is a gold mine for large-scale epidemiological studies.

The inpatient and outpatient registers were used in Studies II and III to identify all infections requiring health care contact occurring from 1997 to 2016 in individuals in the SNMC. In addition, the NPR was used in Studies II and III to identify prevalent conditions as part of the Charlson's comorbidity index (Charlson et al., 1987).

#### **4.2.5 The longitudinal integrated database for health insurance and labour market studies (LISA)**

The longitudinal integrated database for Health Insurance and Labour Market Studies (LISA, Swedish: Longitudinell Integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier) was initiated in 1990 in response of an increasing sick leave level in the country (Ludvigsson et al., 2019). The register contains annual data on all individuals aged 16 or older (aged 15 or older from 2010) registered in Sweden as of December 31 each year from 1990 onwards, and it is administered by Statistics Sweden. Data in LISA derive from several sources including for example the Education Register, the Occupation Register, the Register of Income and Taxation and other administrative sources. Among the information included in LISA are the PIN—through which most linkages are conducted—age, sex, place of residence, civil status, highest level of education, income, sick leave, allowances, disability and retirement pensions, unemployment and other social benefits. From 2001, data on occupational type are also available based on the Swedish Standard Classification of Occupations (SSYK, Swedish: Standard för svensk yrkesklassificering).

The LISA database was used in Study III to identify individuals in the SNMC who were unemployed in 1997. In addition, information on year of retirement was used in Study III to restrict follow-up of individuals who retired after 1997.

# Chapter 5

## Methods

### 5.1 Statistical Methods

#### 5.1.1 The Linear Regression Model

Linear regression (Vittinghoff et al., 2011) is a statistical method commonly used to study continuous outcomes. The main purpose of this model is to analyze how the average value of the outcome  $Y$  changes between groups defined by a given exposure variable  $X$ , eventually controlling for a set of covariates. The model takes the form

$$E(Y|X) = \alpha + \beta X \quad (5.1)$$

or equivalently,

$$Y = \alpha + \beta X + \varepsilon \quad (5.2)$$

where  $\alpha$  is the intercept,  $\beta$  is the slope and  $\varepsilon$  is the error term.  $\alpha$  and  $\beta$  are estimated using ordinary least squares (OLS). The OLS estimates minimize the sum of the squared residuals, namely the sum of the squared differences between the observed outcome and the predicted values provided by a model.

The assumptions of the linear regression model concern the distribution of  $\varepsilon$ . The model assumes that the residuals are independent, normally distributed, have mean 0 and constant variance at all values of  $X$ ,  $\varepsilon_i \sim \text{i.i.d } N(0, \sigma^2)$ . As a means of normalizing the residuals or stabilizing their variance, the natural log transformation of the outcome can be used (Vittinghoff et al., 2011). Despite the assumption of normality concerns the residuals, a reduction of their skewness is often achieved by transforming the outcome. The rationale is that large residuals are usually associated with the more extreme values of the outcome. When the logarithmic transformation is used, a back-transformation of the estimated beta coefficients ( $\exp(\beta)$ ) is interpretable as the expected percentage change in  $Y$  (in the original scale) for a unit increase in  $X$ . Linear regression, with logarithmic transformation of the outcome, was used in Study IV.

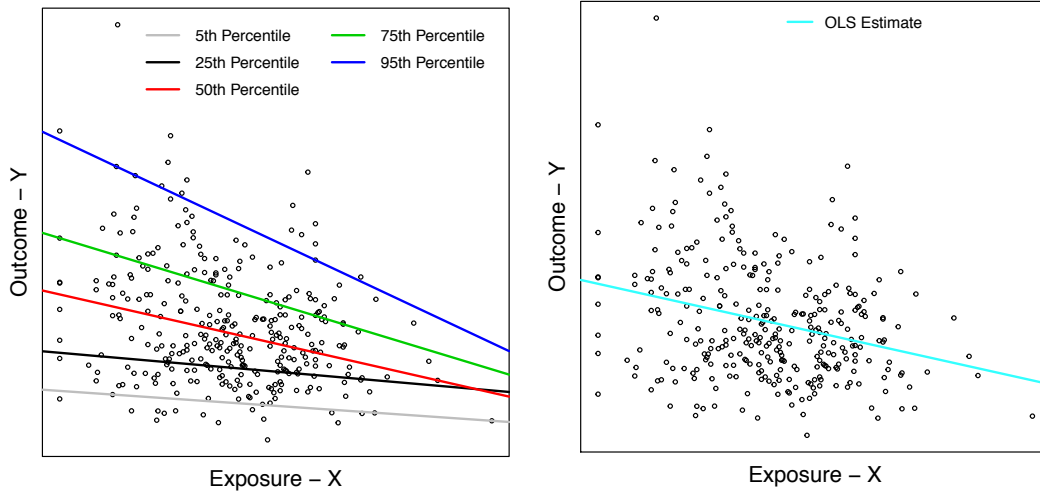
### 5.1.2 The Quantile Regression Model

Quantile regression (Koenker and Hallock, 2001) is a good alternative to the linear regression when the residuals are non-normally distributed or when linearity between outcome and predictors can not be assumed. Quantile regression makes no assumptions on the distribution of the residuals and it provides a full overview of the effect of explanatory variables on the distribution of the outcome. This is possible because quantile regression can quantify relationships across the entire distribution of the dependent variable (e.g. at 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> percentiles), and not only at the mean. The linear model for the conditional  $\tau^{th}$  quantile takes the form

$$Q_{Y_i}(\tau|X) = \alpha_{(\tau)} + \beta_{(\tau)}X \quad (5.3)$$

where  $0 < \tau < 1$ ,  $\alpha_{(\tau)}$  is the intercept for quantile  $\tau$ ,  $\beta_{(\tau)}$  is the slope for quantile  $\tau$

Differently from linear regression, regression coefficients are estimated by minimizing the sum of absolute and not squared residuals. As exemplified in Figure 5.1, quantile regression estimates as many regression models as are the quantiles of interest. In this example, the 5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> percentiles were considered. The effect of X on Y varies across the distribution of the outcome. Whereas almost no effect is observed at lower percentiles, higher percentiles of Y are widely affected by the exposure (negative association). If, for instance, the outcome in Figure 5.1 is BMI, we might conclude that the exposure influences BMI in a different way in underweight and obese individuals.



**Figure 5.1:** Quantile regression models for the association between a continuous exposure X and a continuous outcome Y at various percentiles of Y (on the left). Regression line estimated with Ordinary Least Squares (OLS) (on the right).

The interpretation of quantile regression coefficients is more intuitive than one might think. Just like in linear regression, where the estimated coefficient  $\beta$  tells you how much the mean of Y varies for each-unit increase in X, in quantile regression, the estimated  $\beta$

quantifies the extent to which a certain quantile of the outcome distribution is shifted by a one-unit increase in  $X$ . Each regression line leaves  $(100 - \tau^{th})\%$  of the observations above and  $\tau^{th}\%$  below at any value of  $X$ . For instance, 25% of the observations are above and 75% are below the green line (regression on the 75th percentile) throughout the  $X$  range.

Despite the robustness of quantile regression in respect of outliers and the possibility to obtain much more information about the association of interest, quantile regression is still not widely used in epidemiologic research (Beyerlein, 2014). As a way to encourage its use, we fitted quantile regression models in Study IV.

### 5.1.3 Regression Models for Count Data

Count data are very common in clinical research, examples include e.g. the number of hospitalizations, the number of adverse events occurring during follow-up, and the number of infection recurrences. The main characteristic of count data is that observations can only take non-negative integer values and many zero observations occur; therefore count variables are far from being normally distributed.

#### The Poisson Regression Model

The most commonly used and easy model for count data is the Poisson regression model (Agresti, 2003). The response,  $Y_i$ , is assumed to have a Poisson distribution, which describes the probability that a specific number of events, say  $y_i$ , occurs:

$$Pr(Y_i = y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}; \quad y_i = 0, 1, 2, \dots \quad (5.4)$$

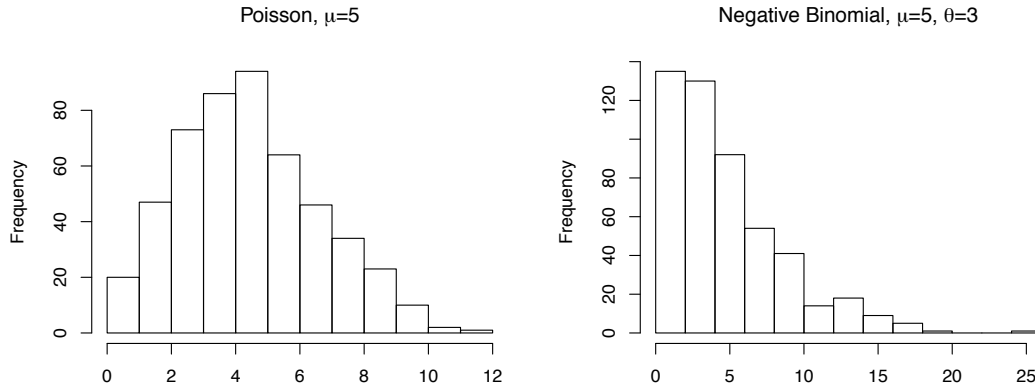
The Poisson model takes the form:

$$\log(\mu) = \log(E(Y|X)) = \alpha + \beta X \quad (5.5)$$

It assumes that the conditional mean and the conditional variance are identical, which is rarely the case in real practice.

#### The Negative Binomial Model

When overdispersion occurs, namely the variance exceeds the mean, a more suitable model is represented by the negative binomial regression (Agresti, 2003). The negative binomial regression belongs, as the Poisson model, to the generalized linear model family. It introduces an extra parameter  $\theta$ , called the dispersion parameter, which allows the mean and the variance to be different. Compared to a Poisson distribution, the negative binomial distribution captures more zeros and it has longer tails (see Figure 5.2). The measure of association between the exposure and the outcome in Poisson regression and negative binomial regression is given by the incidence rate ratio (IRR).



**Figure 5.2:** Simulated data from a Poisson distribution with  $\mu = 5$  and from a negative binomial distribution with  $\mu = 5$  and  $\theta=3$

### The Hurdle Regression Model

Hurdle regression models are well-suited for modelling count data with an excessive proportion of zeros. The intuition behind these models is that there are two distinct processes generating the zeros and the positive values, whereas in standard count data models these two processes are constrained to be the same (Mullahy, 1986).

Differently from the zero-inflated models, where there is a distinction between *structural* and *sampling* zeros, i.e., arising from not-at-risk or at-risk populations respectively, with hurdle regression models all individuals are considered to be at risk and therefore all zeros are considered as *sampling* zeros. Examples of *structural* zeros include number of cigarettes smoked in a certain period among non-smokers (zeros are inevitable); whereas *sampling* zeros arise if current smokers answer that they have not smoked any cigarettes during that specific period, these zeros occur by chance. According to the nature of the outcome being studied, one or the other model is more suitable. A good introduction to these models can be found in the article by Hofstetter, which discusses the main features of these methods and encourages their use (Hofstetter et al., 2016).

The hurdle models consist of two parts: one *zero hurdle* part and one truncated *count* part. The *zero hurdle* part is governed by a binomial probability which discriminates between zero or positive counts. The outcome here is binary,  $Y = 0$  if the count variable is zero,  $Y = 1$  if the count variable is greater than zero, that is, all values larger than 0 are given a value of 1. To estimate the probability of experiencing at least one event, a binomial model or a censored count model can be used. If the realization is positive ( $Y > 0$ ), the threshold (the "hurdle") is crossed and the conditional distribution of positives can be modeled using a truncated count model.

The hurdle model can be expressed as:

$$P(Y_i = y_i | x_i, z_i) = \begin{cases} f_{zero}(0, z_i, \gamma), & \text{if } y_i = 0 \\ (1 - f_{zero}(0, z_i, \gamma)) \frac{f_{count}(y_1, x_i, \beta)}{1 - f_{count}(0, x_i, \beta)}, & \text{if } y_i > 0 \end{cases} \quad (5.6)$$

Where  $z_i$  are predictor variables in the zero part and  $x_i$  represent predictor variables in the count part.

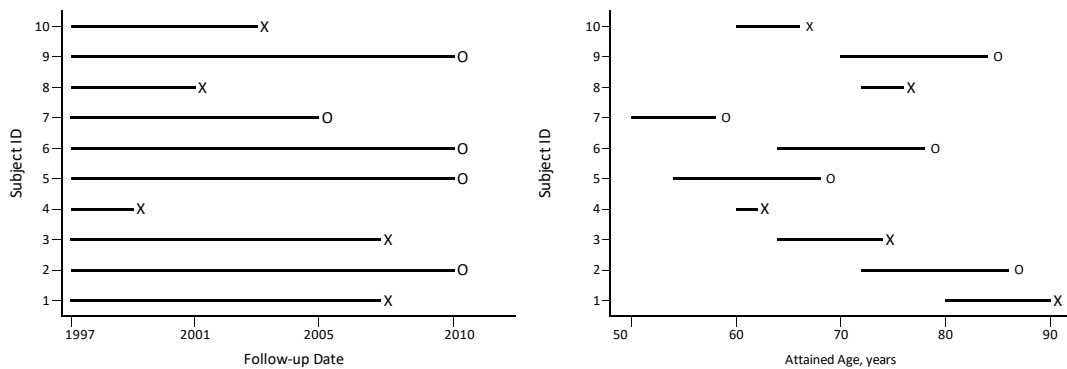
The hurdle regression allows to separately assess the underlying process associated with experiencing, or not experiencing the event of interest and the process associated with the number of events among those experiencing at least one (those who crossed the hurdle). For instance, there might be some predictors influencing the probability of having at least one event, but not being related to the total number of events, and viceversa. The beauty of hurdle regression models is that the zero and the count part can be fitted as two separated models; contrarily to the zero-inflated models where the two classes need to be simultaneously estimated as zeros are modeled in both parts.

Hurdle regression models were used in Study I.

#### 5.1.4 The Cox Proportional Hazard Model

In survival analysis the primary interest is time-to-event occurrence, that is the time from a well-defined time point until the occurrence of a specific event. The main feature of survival data is the presence of censoring (Kleinbaum and Klein, 2012). Censoring occurs when individuals' survival time is unknown and this can either happen because a subject does not experience the event of interest by the end of follow-up, or when a subject drops out of the study for reasons unrelated to the event being studied (e.g. emigration, death for other causes, or unwillingness to continue participating in the study).

Time-on-study is typically used as primary time-scale. In this scenario (represented on the left side of Figure 5.3) all individuals enter the risk set at time 0 (in our example time 0 is 1997), corresponding with e.g. randomization, cancer diagnosis, or filling out a baseline questionnaire.



**Figure 5.3:** Graphical representation of follow-up time for ten individuals included in a study. On the left time-on-study is used as time-scale, whereas on the right the principal time-scale is attained age.

In many observational studies, however, time 0 does not correspond to a meaningful event and therefore individuals are already at risk prior to study entry. It is becoming increasingly common to use attained age as the primary time-scale, especially when age is a much stronger determinant of the outcome than follow-up time (Kom et al., 1997). Additionally,

handling age non-parametrically allows for a more flexible modeling of age. When using attained age instead of time-on-study as the primary time-scale, one should be aware of the delayed entries and left truncation problems. All individuals start to be at risk at their birth date, but they do not come under observation until the start of the study. In this scenario (represented on the right side of Figure 5.3) individuals enter the risk set at the age they complete the baseline questionnaire and exit at their event (or censoring) age.

The Cox Proportional Hazard (PH) Model is the most commonly employed method for analyzing survival data (Cox, 1972). The Cox PH model is a semi-parametric model and has the form:

$$\lambda_i(t) = \lambda_0(t) \exp(X_i \beta) \quad (5.7)$$

Where  $t$  is time,  $\lambda_0(t)$  is the baseline hazard,  $X_i$  represents covariates, and  $\beta$  their estimated coefficients.

The PH Cox model focuses on the hazard  $\lambda(t)$ , which is the conditional probability of the event of interest to occur in the interval  $[t, t + dt)$ , given that it has not occurred before. The measure of association between the exposure and the outcome is given by the hazard ratio (HR), defined as the hazard in the exposed, divided by the hazard in the unexposed.

The model assumes that the effects of the predictor variables upon survival are constant over time. The hazards may vary over time, but their ratio is the same at all time points. The underlying PH assumption can be tested with graphical and more formal methods (e.g., using Schoenfeld residuals test) and extensions of the traditional Cox model are available in case of departures from proportionality. Stratified Cox models, for instance, are a valid option for covariates which are not of primary interest for the analysis. The covariate displaying non-proportionality is fitted without constraints of proportionality and different baseline hazards are estimated for each level of the covariate.

### Parametric Survival Analysis

Weibull regression model is one of the most common parametric regression models for survival analysis. Contrarily to Cox regression, where the baseline hazard is not specified, the Weibull model assumes the following:

$$\lambda(t) = \lambda_p t^{p-1} \quad (5.8)$$

Where

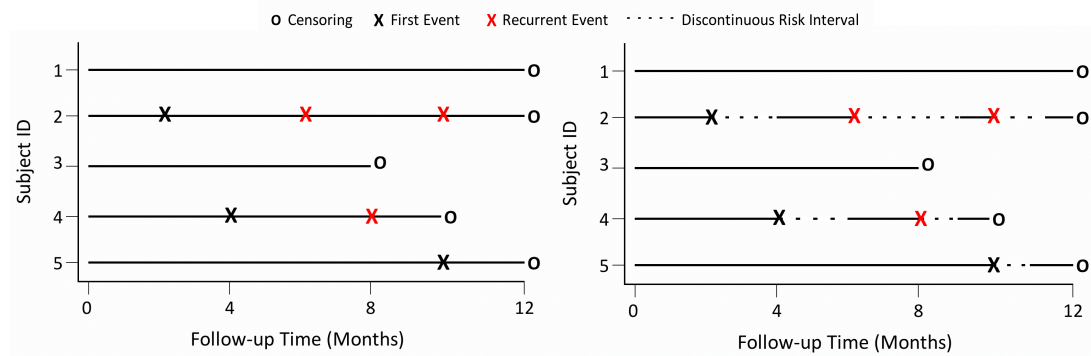
$$\lambda = \exp(\beta_0 + \beta_1 X) \quad (5.9)$$

and  $p$  is the shape parameter. If  $p > 1$  the hazard increases, if  $p = 1$  the hazard is constant, if  $p < 1$  the hazard decreases over time. If reasons exist to assume that the baseline hazard follows a particular form, parametric models provide greater efficiency compared to semi-parametric models as they estimate less parameters.



### 5.1.5 Recurrent Event Survival Analysis

So far, we assumed that the outcome under investigation can occur only once for each individual. However, in many research studies in which the event of interest is not fatal (i.e. death), a subject may experience an event multiple times over follow-up (e.g. admissions to hospital, cardiovascular events, infections, cancer recurrences). Figure 5.4 displays the observational time of five subjects in the presence of recurrent events. The main difference compared to the traditional approach is that the follow-up does not end at the time of the first event and each individual is observed until the end of the study, or until his/her drop-out date.



**Figure 5.4:** Graphical representation of follow-up time for five individuals included in a study where repeated events are allowed to occur. On the left is represented a scenario without discontinuity of risk intervals, whereas on the right discontinuous risk intervals are present.

A special case of recurrent events occurs when risk intervals are discontinuous (see Figure 5.4 on the right). This happens when individuals are not at risk of a second event if they haven't recovered from the previous one (Guo et al., 2008). For example, if the outcome of interest is any hospitalization for infections, individuals can not be at risk of a second hospitalization until their discharge date. As a consequence, the duration of the previous event has to be taken into account in the time at risk count. This discontinuous time is represented by the dotted line in Figure 5.4.

Despite that several modeling techniques are available for analysis of recurrent events (Amorim and Cai, 2015; Westbury et al., 2016; Ozga et al., 2018), most often events subsequent the first are ignored. Using Poisson regression models or negative binomial regression is still an option, but they ignore the time between repeated occurrences, which leads to an inefficient use of data. The main characteristic of recurrent event data is that failure times are correlated within the same subject and since the independence assumption of the traditional cox model is violated, statistical methods that take into account the lack of independence are needed.

### The Andersen-Gill model (AG)

The AG model (Andersen and Gill, 1982) is a counting process model which assumes independence between all event times, regardless if they belong to the same individual or to different individuals. It assumes that all failures are identical and the order is unimportant, therefore the risk of experiencing an event at time  $t$  does not change according to the number of previous events. This is a simple extension of the traditional Cox model, with the only difference that individuals remain in the risk set after the occurrence of the first event. If the assumption of independent recurrent event times is not fulfilled, time-dependent covariates can be used to capture the dependencies. Otherwise, a robust sandwich covariance matrix can be used to account for correlation.

The AG model is suitable when interest is in the overall effect of a predictor on the intensity of the occurrence of an event and when it is reasonable to assume that the risk remains the same, regardless whether previous events occurred or not. Additionally, it is a good choice when the correlation among events is induced by measured covariates.

### The Prentice, Williams and Peterson Model (PWP)

The PWP model (Prentice et al., 1981) handles ordered recurrent events by stratification. Contrarily to the AG model, it does not assume a common baseline hazard function for each event, but instead the hazard of a recurrent event is influenced by the preceding event history. Also, the PWP model allows to estimate strata-specific effects introducing an interaction term between the predictor variable and the variable identifying the event number.

The form of the model is similar to the traditional Cox model, with the addition that it estimates a separate hazard function for each repeated event  $k$ :

$$\lambda_{ik}(t) = \lambda_{0k}(t) \exp(X_{ik}\beta) \quad (5.10)$$

Where  $\lambda_{ik}(t)$  is the hazard function for the  $k^{th}$  event of the  $i^{th}$  subject.

Two different formulations of the model are available. The first evaluates the effect of a covariate for the  $k^{th}$  event since study entry (total time scale), whereas the other consider the time since the previous event (gap time scale).

Risk sets are defined distinctly for each event stratum. All individuals are at risk for the first event, but only those that experience a previous  $(k - 1)^{th}$  event are at risk for the successive one ( $k^{th}$ ). It might happen that, when few individuals have a large number of events, the risk-sets become very small as the number of strata increases. It is a common practice to restrict the maximal number of events for individual subjects to avoid small risk-set in later strata.

The PWP model is preferred when there are few recurrent events per subject and when it is appropriate to assume that the occurrence of an event is influenced by events in the

past. However, when there is no strong biological relationship between subsequent events, the PWP model could considerably underestimate the overall effect (Guo et al., 2008).

### Frailty Models

Frailty models are random effects models which account for heterogeneity by introducing a random component in the model (Kleinbaum and Klein, 2012). Frailty models can be used in the framework of recurrent events as the random component introduced can account for the dependence observed among the recurrent events. The model has the form:

$$\lambda_i(t) = \lambda_0(t) \alpha_i \exp(X_i\beta) \quad (5.11)$$

Where  $\alpha_i$  is the random component that describes the excess risk (or frailty) for each individual.

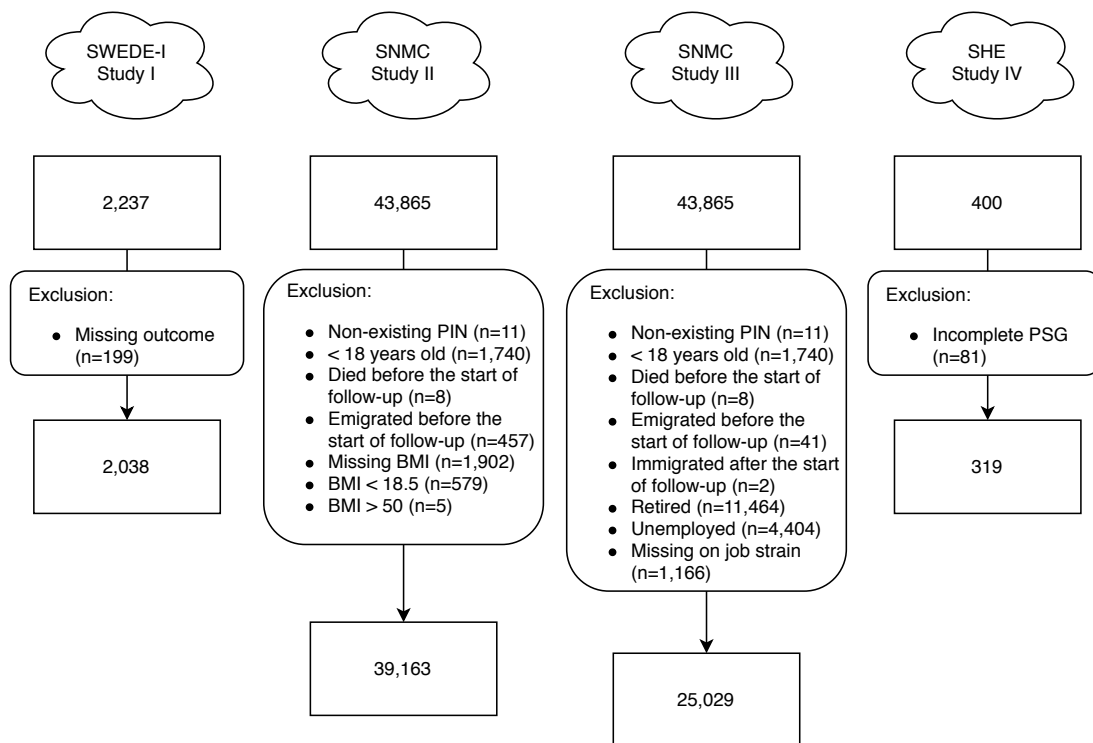
Frailty are assumed to be individual-specific and are constant over time within subjects. Most commonly, the random component is assumed to follow a gamma distribution. It not only has an impact on standard errors, but also on estimated coefficients which can be very different from AG and PWP models if the random term plays an important role (Guo et al., 2008). Frailty models assume that the baseline hazard does not vary by event. Furthermore, order of events is not taken into consideration. Frailty models are appropriate when the susceptibility to the risk of recurrent events is heterogeneous across individuals.

To summarize, the choice of the most suitable approach is not straightforward and it is determined by many factors, including number and order of events, relationship between consecutive events, biological mechanisms and research question. Survival models for recurrent events were used in studies II and III.

#### 5.1.6 Principal Component Analysis

Principal component analysis (PCA) (Rencher and Christensen, 2012) is a statistical technique for reducing the dimensionality of the data. PCA uses orthogonal transformations to find a new set of uncorrelated variables (principal components) which are linear combination of the observed variables. Principal components are chosen in such a way that they capture most of the original variability. Most often, to facilitate the interpretation of each principal component, rotated loadings—the weight by which each standardized original variable should be multiplied to get the component score—are preferred. PCA was used in Study IV to reduce the number of sleep variables derived from sleep questionnaires and PSG.

## 5.2 Overview of Methods by Study



**Figure 5.5:** Detailed exclusion criteria applied to studies included in the thesis.

### 5.2.1 Study I

#### Study Population

Study I assessed the association between physical activity and sleep habits with the incidence of upper respiratory tract infections (URTI) in a prospective cohort of employed individuals living in Eskilstuna. Participants in the study were those included in the SWEDE-I cohort with non-missing information on the outcome. The final sample size consisted of 2,038 workers (Figure 5.5).

#### Study Design

Study I has a prospective design. Over the course of follow-up, participants completed five questionnaires which allowed to collect enormous quantity of data on different domains. As the big effort required to answer to all questions, the first four questionnaires were sent out monthly one-by-one in a paper version. Nevertheless, they were all immediately available when the study started for those deciding to answer on the web. The fifth questionnaire was available only in an electronic form, as it contained a web-based physical activity

questionnaire previously validated (Bonn et al., 2012). Participants were followed-up for 9 months (between September and May) for any occurrence of respiratory infections.

## **Exposure and Outcome measures**

### *Exposure*

Information on sleep habits were obtained from two multiple-choice questions in the fourth questionnaire: "How many hours, approximately, do you usually sleep on an ordinary week-day?" and "How do you usually sleep?" Only 1% of individuals reported to sleep  $\geq 9$  hours and therefore long sleep was considered as sleeping 8 hours or more.

Physical activity was measured as the total energy expenditure during an ordinary 24-hour day and it was expressed as metabolic energy turnover hours per day (MET-h/d). The daily energy expenditure was computed from three dimensions: work time, leisure time, and time spent sleeping. Two different instruments were used to collect information on time spent at different intensity levels during work and leisure time. The web questionnaire Active-Q (Bonn et al., 2012) was used to measure leisure time physical activity, whereas activities at work were assessed using a modified version of a validated physical activity instrument for self-reported total physical activity (Lagerros et al., 2006). Sleeping hours were assigned a MET value of 0.9. These three distinct dimensions were summed up to obtain the total MET-h/d, which was then categorized into tertiles for the analyses.

### *Outcome*

During the entire follow-up, individuals self-reported any onset of fever, respiratory infections and gastroenteritis. At each event they responded to a symptom questionnaire on the web or through an automated voice response. Additionally, participants were instructed to take a nasal swab at each event occurrence.

Infections were defined as the presence of a sick report or a nasal swab, irrespectively of the positiveness of the test. An infection was classified as URTI if individuals reported at least one of the following symptoms: cough, runny nose, or sore throat. Individuals with infections accompanied by symptoms of a different origin, or individuals with no sick reports and no nasal swabs were considered as not having experienced any URTI during follow-up. URTI were classified as missing when infections occurred, but, due to missing answers to all three URTI symptoms, the distinction between URTI or other infections was not possible.

## **Statistical Analysis**

Due to the high proportion of individuals with zero infections during the follow-up, which led to a highly skewed distribution of the outcome, the associations of interest were investigated using hurdle regression models. The positive counts part was modeled using

a left-truncated negative binomial regression; whereas the binary zero-hurdle part was modeled using a right-censored Poisson regression. Missing values on incomplete variables were imputed before fitting the hurdle regression models.

### 5.2.2 Study II

#### Study Population

Study II examined the association between obesity and risk of infections using data from the SNMC. The study population included in this study is presented in the flowchart in Figure 5.5. After excluding individuals who were not traceable in the registers due to a non-existing PIN, individuals younger than 18 years of age, individuals who died or emigrated before the start of follow-up, and individuals with missing information on height or weight, underweight individuals and those with BMI above 50 (99.99th percentile), the final cohort consisted of 39,163 man and women.

Three individuals died between the fundraising event (which took place September 10-14, 1997) and October 1, 1997 which corresponds to the beginning of follow-up. Five individuals had a registered date of death antecedent the time of the National March, which is probably due to failures in reporting/decoding the PIN on the questionnaire filled in by hand. The high number of individuals emigrated outside Sweden before the start of follow-up is mainly due to two reasons. First, some of the emigrated individuals might have moved to close Nordic countries and decided to participate in this nationwide event, despite not being Swedish residents anymore; others might have been de-registered from the TPR due to some years spent abroad before September 1997, and be back in Sweden at the time of the National March. As we had not received information from Statistics Sweden regarding re-immigration dates, those individuals were considered as being still abroad.

#### Study Design

Study II is a prospective cohort study. Participants filled in a questionnaire at baseline and were followed-up through record linkages until first infection (of any kind, or type-specific accordingly to the analysis being conducted), death, emigration, or end of follow-up (December 31, 2016), whichever occurred first. In a secondary analysis, where all infections—and not only the first—occurring during follow-up were taken into consideration, follow-up ended at the time of emigration, death or on December 31, 2016.

#### Exposure and Outcome measures

##### *Exposure*

BMI was the main exposure of interest in this study. Individuals self-reported height and weight in the baseline questionnaire; BMI was then computed as weight in kilograms divided by height in meters squared. Based on the standard classification from WHO, in-

dividuals were classified into normal weight (18.5–24.9 kg/m<sup>2</sup>; reference), overweight (25–29.9 kg/m<sup>2</sup>) and obese ( $\geq 30$  kg/m<sup>2</sup>) (WHO, 2000).

Secondary measures of adiposity considered were waist circumference and waist-to-hip ratio. Based on the WHO sex-specific cut-off points for Caucasians (identified based on a high and very high risk for cardiovascular disease and diabetes type 2), waist circumference thresholds were set at >80 cm and >88 cm for women and >94 cm and >102 cm for men (WHO, 2000). According to WHO guidelines, abdominal obesity was defined as WHR >0.85 for women and >0.90 for men (WHO, 2011).

### *Outcome*

Infections requiring health care contact were examined as outcome in this study. Infections were ascertained from the NPR using the ICD-10 codes (Supplementary Table 1, Paper II) and classified as upper respiratory tract infections, lower respiratory tract infections, gastrointestinal tract infections, skin and subcutaneous tissue infections, urinary tract infections, gynaecological infections, sepsis, and other infections. First, we considered infections from both the inpatient and the outpatient registers as events of interest; then, secondly, only infections requiring overnight hospitalization were registered as event.

## **Statistical Analysis**

### *Time-to-first Infection*

Cox PH models with age as the underlying time-scale were fitted to estimate hazard ratios (HRs) and 95% confidence intervals (95% CI) for infection occurrence, both overall and infection-specific. The analyses were adjusted for the following potential confounders, which were selected based on subject matter knowledge: highest attained educational level, physical activity, smoking status, alcohol consumption, sleep duration, disposable income, Charlson's comorbidity index at baseline and, in women only, for hormone treatment. Charlson's index was the only predictor not fulfilling the Cox PH assumption, therefore Stratified Cox models on this comorbidity score were used throughout the analyses. To assess linear trends we created a new categorical variable based on the median of each category of BMI. This new variable was then entered into the models as a continuous variable and tested for statistical significance.

### *Multiple-failure analysis*

Extensions of the traditional Cox PH model—namely the AG model, the PWP model, and frailty models—were applied to study the association between BMI and all infections occurring during follow-up. The maximum number of infections for each individual was restricted to seven in order to prevent small risk-sets. As for frailty models, we fitted a Weibull regression model with a shared frailty term assumed to follow a gamma distribution.

### 5.2.3 Study III

#### Study Population

Study III explored the association between job stress and risk of infections in individuals in the SNMC. The source population of this study is the same as in Study II, but different criteria of exclusion were applied (Figure 5.5). Beside excluding individuals with non-existing PIN, under 18 years old, emigrated or died before the start of follow-up, we also excluded those who were retired or unemployed in 1997 and those with missing answers on at least half of the items used to compute the job demand and/or job control scores. The number of individuals excluded since they were not registered in Sweden in 1997, is much smaller compared to the number reported in Study II as we received updated information from Statistics Sweden for this study. The new linkage contained not only emigrations, but also re-immigration dates so that we could figure out who were the individuals who came back to Sweden after residing abroad and therefore present in the TPR in 1997. In addition, according to the Migration Register, two individuals were registered in the TPR for the first time a few years after the baseline questionnaire. They were excluded as we could not guarantee the accuracy of their follow-up through record-linkages.

#### Study Design

Study III employed a prospective cohort design using time-to-event data. Participants, followed-up through record linkages, were censored at retirement, emigration, death or end of follow-up.

#### Exposure and Outcome measures

##### *Exposure*

**Table 5.1:** Items included in the Job Demand and Job Control scores

<b>Job Demand<sup>a</sup></b>	<b>Job Control<sup>a</sup></b>
Work very fast	Learn new things
Work very hard	Non-repetitive work
Too much demand	Creativity required
Not enough time	High skill level
	Decide on how you do your work
	Decide on what you do in your work

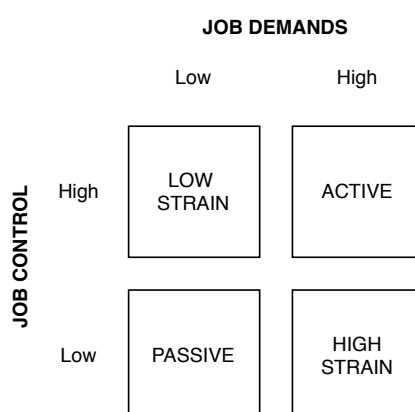
<sup>a</sup> Responses were given on a 4-point scale ranging from 1=Sel-dom/Never to 4=Always/Almost always.

Stress at work was measured using the Swedish version of the Demand-Control Questionnaire (DCQ) (Theorell et al., 1988). As implicit in its name, the questionnaire covers two different dimensions of the work environment: one dimension, named *Job Demand*, is related to the workload and time available to carry out working tasks; the second dimension,



called *Job Control*, concerns the decision latitude and skills. Specific items defining each of these two dimensions are reported in Table 5.1.

Scores of Job Demand and Job Control, computed as the mean of answers given in items in Table 5.1, were categorized into tertiles and used as the main exposures in the analyses. Job strain was then defined following the quadrant approach (Figure 5.6), first introduced by Karasek in 1979 (Karasek Jr, 1979).



**Figure 5.6:** The Job Strain Model. "High" is defined as having a score above the median, "Low" as having a score below the median.

### Outcome

Infections were defined using the same ICD-10 codes as in study II. In this study, however, only the following subgroups were considered: upper respiratory tract infections, lower respiratory tract infections, gastrointestinal tract infections, skin and subcutaneous tissue infections, urinary tract infections and other infections. No distinction was made between infections diagnosed in inpatient or outpatient care.

### Statistical Analysis

As in Study II, both time-to-first infection and multiple failure models were implemented in the analysis of survival data. To be more specific, the AG model was used when considering infections overall, whereas the PWP model was used when considering only infections belonging to the same subgroup. All models being fitted used age as the underlying time-scale. The PH assumption was tested using Schoenfeld residuals and stratified Cox models were used to adjust for variables not fulfilling the PH assumption (sex).

Based on data on occupation available in the LISA database from 2001, we could classify individuals into white- and blue-collar (identified using the SSYK) and conduct subgroup analyses by occupational types. White collars are those whose jobs require university or higher education; whereas blue collars usually perform manual labor.

The assumptions on the relationships between variables were represented in a directed acyclic graph (DAG) (Textor et al., 2016). This tool was used to select the minimal set of

variables sufficient to obtain unbiased estimates of the association of interest. In a first step the models were adjusted for sex, education, physical activity, shift work, disposable income and Charlson's comorbidity index. Analyses were then repeated by adding the potential intermediate variables in the models: BMI, smoking status, sleep duration and alcohol consumption (results shown in Tables 2,3, and 4 in Manuscript III).

#### 5.2.4 Study IV

##### Study Population

Study IV investigated the relationship between sleep characteristics and markers of inflammation in women included in the SHE cohort and who were randomly selected to participate in Phase II of the SHE study. Exclusion criteria are summarized in Figure 5.5. Only women with incomplete PSG were excluded from the analyses.

##### Study Design

Study IV is a cross-sectional study. Women filled in a sleep questionnaire in the evening, underwent in-home overnight PSG, and had blood tests done the morning after the PSG.

##### Exposure and Outcome measures

###### *Exposure*

Sleep characteristics were measured both subjectively through the Uppsala Sleep Inventory (Hetta et al., 1999) and objectively through PSG. Main variables collected are shown in Table 5.2. Principal components retrieved from PCA were used as the main exposures of interest.

**Table 5.2:** Sleep characteristics measured with the Uppsala Sleep Inventory and from in-home overnight polysomnography.

Sleep Questionnaire	Polysomnography
Total sleep time, min	Total sleep time, min
Sleep quality, 0-100	Sleep efficiency, %
Epworth sleepiness scale	Sleep latency, min
Difficulties initiating sleep <sup>a,b</sup>	REM latency, min
Difficulties maintaining sleep <sup>a,b</sup>	N3, %
Early morning awakening <sup>a,b</sup>	N3, min
Feeling tired <sup>a,b</sup>	REM, %
Feeling sleepy	REM, min
Non-resting sleep <sup>a</sup>	Apnea-hypopnea index
Difficulties in memory <sup>a</sup>	Arousals
Difficulties to concentrate <sup>a</sup>	Awakenings
	Stage changes

<sup>a</sup> Measured on a scale ranging from 1 to 5 (the higher the worst)

<sup>b</sup> Items used to define the presence of insomnia symptoms.

PSG is considered the gold standard to measure sleep objectively. PSG is conducted during nights when the individuals are asleep and it allows to collect several physiological parameters, including brain electrical activity through electroencephalography (EEG), muscle tone through electromyography (EMG), eye movements through electrooculography (EOG), heart rhythm through electrocardiography (ECG) and several other respiratory parameters. In-home PSG is an emerging method to measure sleep parameters. It is more representative of usual sleep and less affected by the first-night effect typical of laboratory PSG (Edinger et al., 1997).

#### *Outcome*

The outcomes in this study are markers of inflammation. Specifically, the main interest was into CRP, followed by IL-6 and TNF $\alpha$ . Normal values of CRP were defined as having CRP less than 5 mg/L (WHO, 2014a).

#### **Statistical Analysis**

As first step, PCA was used to reduce the dimensionality of the data and retrieved principal components were entered in the subsequent models as predictors of interest. Due to a positive skewed distribution of CRP, IL-6 and TNF $\alpha$ , we fitted linear regression models with logarithmic transformation of the outcomes. The association between sleep and inflammation was also examined using quantile regression.

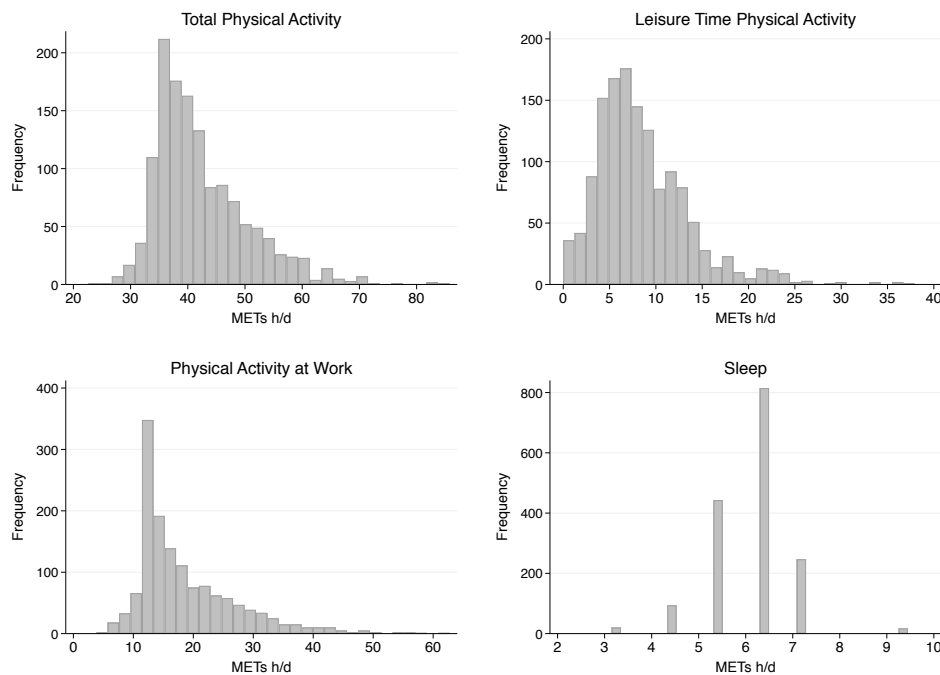
# Chapter 6

## Main Results

### 6.1 Habitual sleep, Physical activity and URTI (Study I)

Among the study participants, 7% reported sleeping for 5 hours or less per night, 27% reported sleeping for 6 hours, 50% for 7 and 16% for 8 hours or more. The majority of participants reported a good or quite good quality of sleep (75%).

The distribution of total physical activity, measured as total MET-h/d, is displayed in figure 6.1 together with the distributions of MET-h/d for each physical activity component: leisure time, work and sleep. The peak in work physical activity was observed at 12 MET h/d and it belonged to individuals who answered that they were involved in activities such as office work, sitting in a meeting (requiring 1.5 MET h/d) for most of their working hours ( $8 \times 1.5 = 12$  MET h/d). The peak in MET h/d related to sleeping time is observed in



**Figure 6.1:** Distribution of total physical activity, leisure time physical activity, physical activity at work and METs h/d during sleep in employed individuals in the SWEDE-I cohort.

correspondence of 6.3 MET h/d and it belongs to individuals sleeping 7 hours per night ( $7 \times 0.9 = 6.3$  MET h/d). As physical activity was categorized into tertiles, individuals were equally distributed across the three groups of physical activity identified by the following cut-offs: 37.7 MET-h/d and 43.9 MET-h/d.

A total of 1,597 URTI occurred in 1,583 person-years, resulting in an incidence of 1.01 infections/person-year (95% CI 0.96-1.06). The distribution of URTI presented an excess of zeros, as half of the participants did not report any URTI event. The total number of URTI ranged from 0 to 8, with 31% of individuals reporting one event, and 19% reporting two events or more.

Contrarily to our hypotheses, neither sleep duration nor sleep quality were related to the occurrence or to the number of URTI (see Table 6.1). In addition, we did not find any interaction between these two sleep dimensions. Similarly, physical activity did not show any inverse association with URTI events, neither in the zero-hurdle part, nor in the positive counts one (see Table 6.1).

**Table 6.1:** Adjusted Incidence Rate Ratios (IRR) and 95% CI of the relation between sleep habits, physical activity and URTI, fitted in a hurdle regression model.

Exposure	Zero-hurdle			Positive Counts		
	IRR <sup>a</sup>	95% CI	p-value	IRR <sup>a</sup>	95% CI	p-value
<b>Sleep duration (hours)</b>			0.37			0.29
≤5	0.96	0.72-1.29		0.79	0.51-1.20	
6-7	1.00	Reference		1.00	Reference	
≥8	1.14	0.94-1.38		0.86	0.67-1.11	
<b>Sleep quality</b>			0.08			0.26
Quite good/Good	1.00	Reference		1.00	Reference	
Neither bad nor good/Bad	1.16	0.98-1.37		1.13	0.92-1.39	
<b>Physical activity (MET-h/d)</b>			1.00			0.46
<37.7	1.00	Reference		1.00	Reference	
37.7-43.9	1.00	0.84-1.20		1.05	0.83-1.31	
≥43.9	1.01	0.84-1.21		0.91	0.71-1.15	

<sup>a</sup> IRR, Incidence Rate Ratio adjusted for age, sex, presence of children ≤6 years old in the household, smoking, travel mode to work, education, BMI, number of close contacts per day, treatment for allergy, asthma or lung cancer, and treatment for immunodeficiency or transplantation.

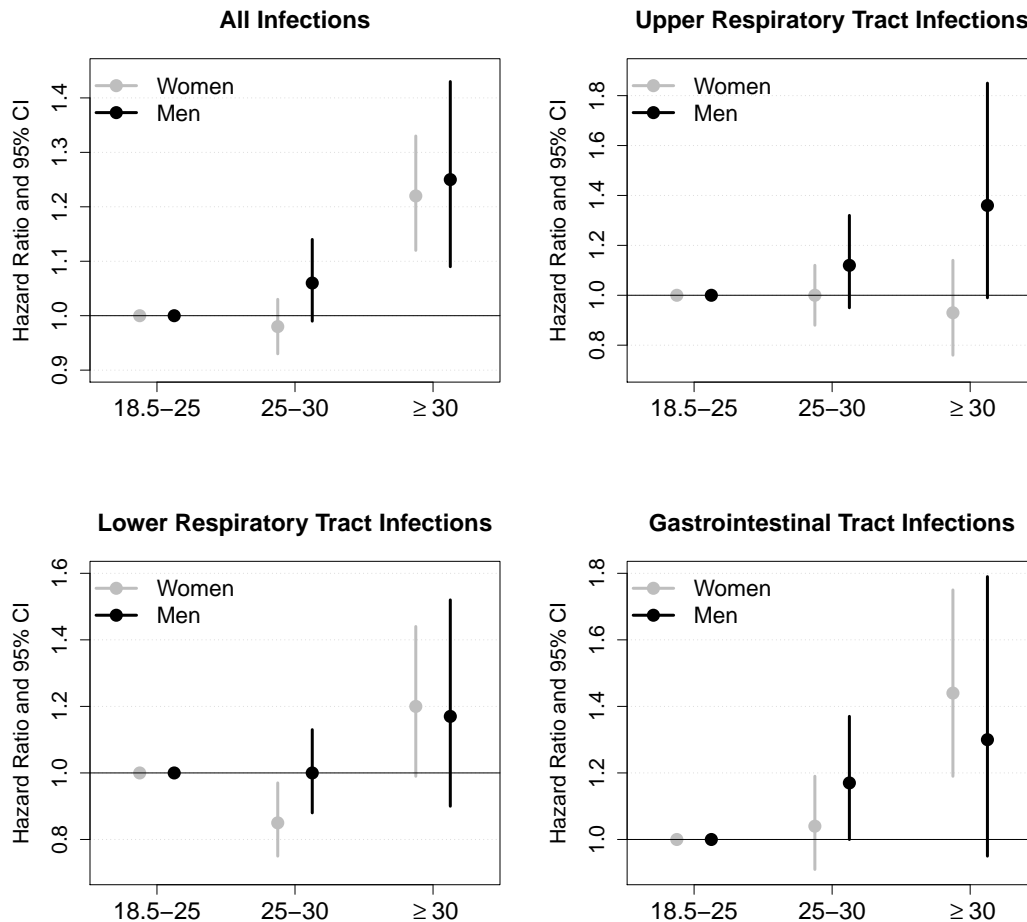
The chance of experiencing URTI was higher in females, in individuals with children below 6 years in the household, with higher education and receiving treatment for allergy, asthma or lung cancer. Factors associated with URTI count among those with at least one reported event were having children below 6 years and female gender (see Supplementary Files S3 and S4 in Paper I).

## 6.2 Obesity and Infections (Study II)

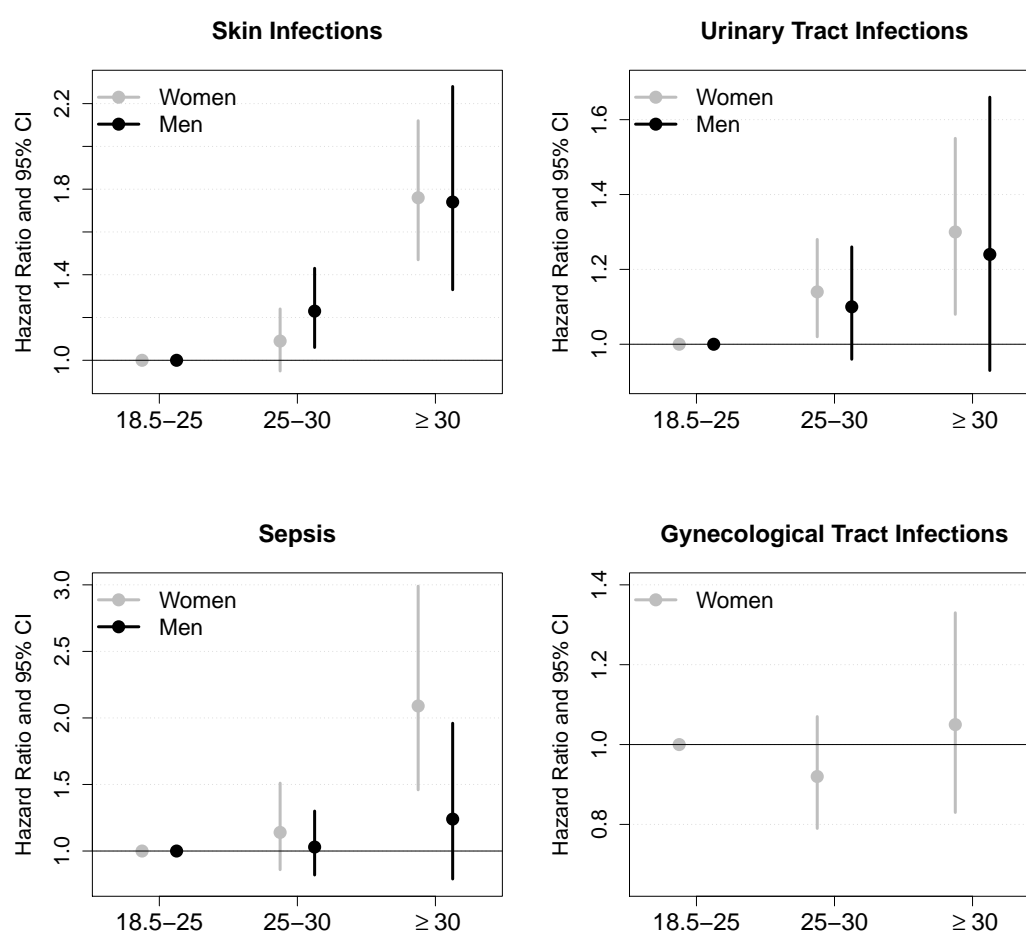
Globally, rates of obesity are higher in women than in men. In the SNMC, 29% of women and 41% of men were overweight, and 8% of women and 6% of men were obese.

During a mean follow-up of 15.4 years, 27,675 infections requiring healthcare contact were observed, with an average of 2.2 infections per subject and 32% of individuals with at least one event. The number of infections ranged from 0 to 47, but 99% of participants experienced at most 7 infections. The majority of the infections (62%) occurred in outpatient care. Upper respiratory tract infections were the most common infections among women, whereas lower respiratory tract infections were the most frequent in men.

In the analysis of time-to-first infection we found an increased incidence of any infection in women (HR=1.22; 95% CI=1.12; 1.33) and men (HR=1.25; 95% CI=1.09; 1.43) with obesity compared to normal weight individuals. In particular, in subjects with obesity, we found a higher incidence of skin infections (in both genders), gastrointestinal tract infections (in women), urinary tract infections (in women), and sepsis (in women). Results are summarized in Figure 6.2 and further details can be found in Table 2 in Paper II.



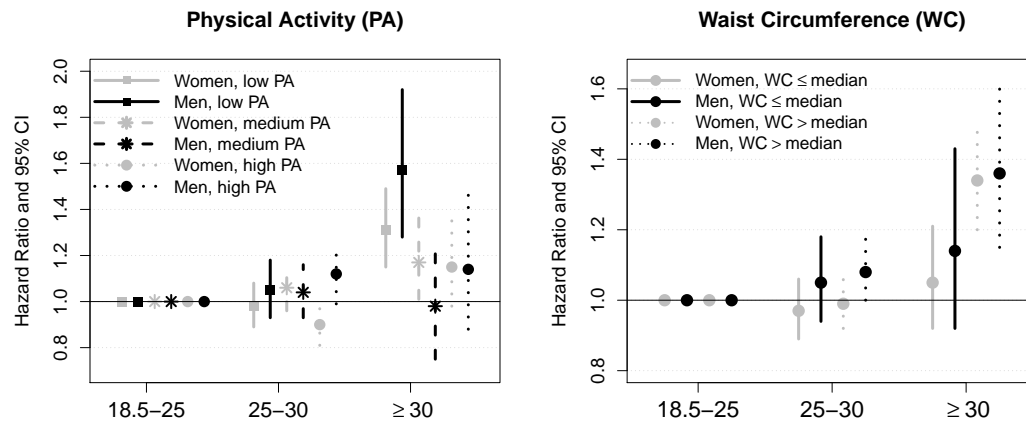
**Figure 6.2:** Adjusted Hazard Ratios and 95% Confidence Intervals of the association between BMI categories and infection incidence, overall and infection specific, using traditional PH Cox models. Results are stratified by gender and being normal weight was taken as the reference.



**Figure 6.2:** (continued) Adjusted Hazard Ratios and 95% Confidence Intervals of the association between BMI categories and infection incidence, overall and infection specific, using traditional PH Cox models. Results are stratified by gender and being normal weight was taken as the reference.

As we found a multiplicative interaction between BMI and physical activity, and BMI and waist circumference, we replicated the analyses stratified by levels of these two variables (see Figure 6.3). The association between obesity and infections was stronger in individuals with low levels of physical activity and in those with a waist circumference above the median. On the other hand, obese individuals physically active or with a waist circumference below the median did not have an increased risk of infections.

When accounting for repeated events, we found similar estimates as the traditional Cox models. In particular, the AG model returned estimates really close to the aforementioned ones (HR=1.26; 95% CI=1.15; 1.39 for women; HR=1.24; 95% CI=1.07; 1.44 for men). The PWP models were the most conservative (HR=1.17; 95% CI=1.10; 1.26 for women; HR=1.17; 95% CI=1.05; 1.30 for men), whereas frailty models revealed an association between obesity and infections in women only (HR=1.17; 95% CI=1.05; 1.30 for women; HR=1.04; 95% CI=0.88; 1.22 for men).



**Figure 6.3:** Adjusted Hazard Ratios and 95% Confidence Intervals of the association between BMI categories and infection incidence using traditional PH Cox models. Results are stratified by gender and physical activity (on the left) and by gender and waist circumference (on the right). Being normal weight was taken as the reference.

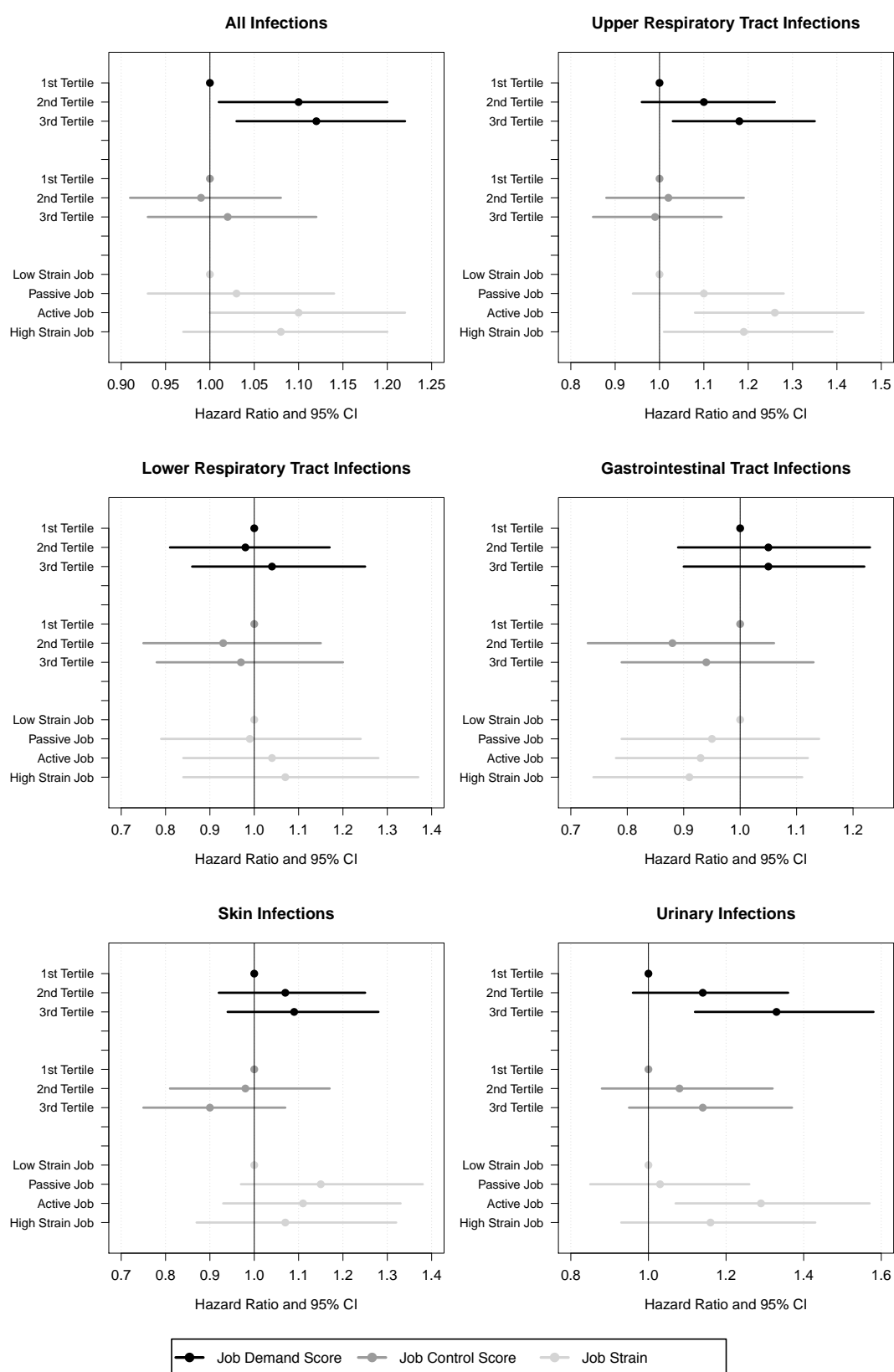
### 6.3 Job Stress and Infections (Study III)

The individuals included in the study were distributed as follows: 23% with low strain jobs, 30% with passive jobs, 25% with active jobs and 22% with high strain jobs. During a median follow-up of 16.2 years, 8,257 infections requiring health care contact were observed, with an average of 1.8 infections per subject and 18% of individuals with at least one event.

The incidence of infections was higher in individuals in the third tertile of job demand score, compared to individuals in the first tertile (HR=1.12; 95% CI=1.03; 1.22). Specifically, high job demands were associated with higher incidence of upper respiratory tract infections (HR=1.18; 95% CI=1.03; 1.35) and urinary tract infections (HR=1.33; 95% CI=1.12; 1.58), but we did not find evidence for an association with lower respiratory tract infections, gastrointestinal tract infections or skin infections (see Figure 6.4). On the other hand, job control did not exhibit any inverse association with infection incidence, neither overall, nor infection specific. As for job strain, high strain jobs did not differ statistically compared to low strain jobs in respect of infection occurrence; only active jobs showed an increased risk of infections compared to the reference group, in particular upper respiratory tract infections (HR=1.26; 95% CI=1.08; 1.46) and urinary tract infections (HR=1.29; 95% CI=1.07; 1.57).

Further adjustment for potential intermediate variables did not induce a marked change in the estimates presented here. Further details can be found in Manuscript III. The increased risk of infections of any type found in high demand jobs, and the absence of association with job control, were confirmed from analyses on time-to-first infection presented as Supplementary Tables S2, S3, and S4 in Manuscript III.



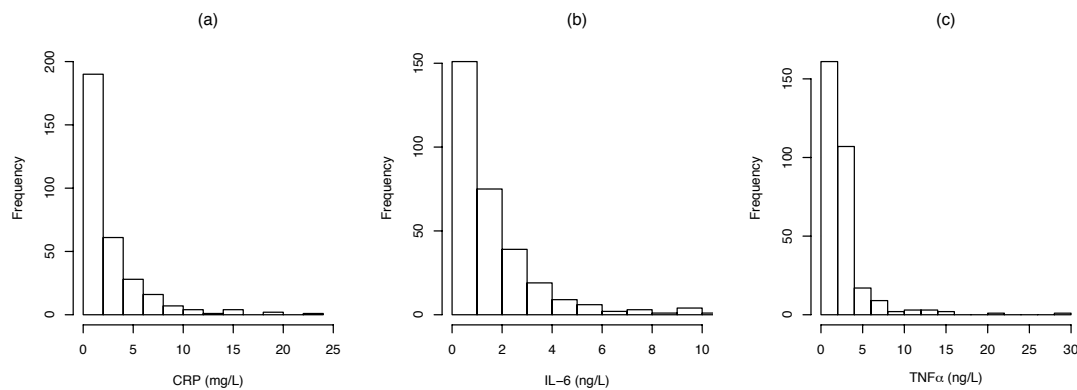


**Figure 6.4:** Adjusted Hazard Ratios and 95% Confidence Intervals of the association between Job Demand, Job Control and Job Strain with infections occurrence using PH Cox models for repeated events. The first tertiles of Job Demand and of Job Control and the Low Strain Job category were taken as the references.

## 6.4 Sleep and Inflammation (Study IV)

In this study 30% of women reported to have at least one insomnia symptom (difficulties initiating sleep, difficulties maintaining sleep, or early morning awakenings), and 11%, beside having insomnia symptoms, also felt tired during the day, which is compatible with an insomnia diagnosis. According to PSG, women slept on average 6.26 hours, with 17% of time spent in REM sleep. The sleep efficiency was 83% with an average of 20 minutes of sleep latency.

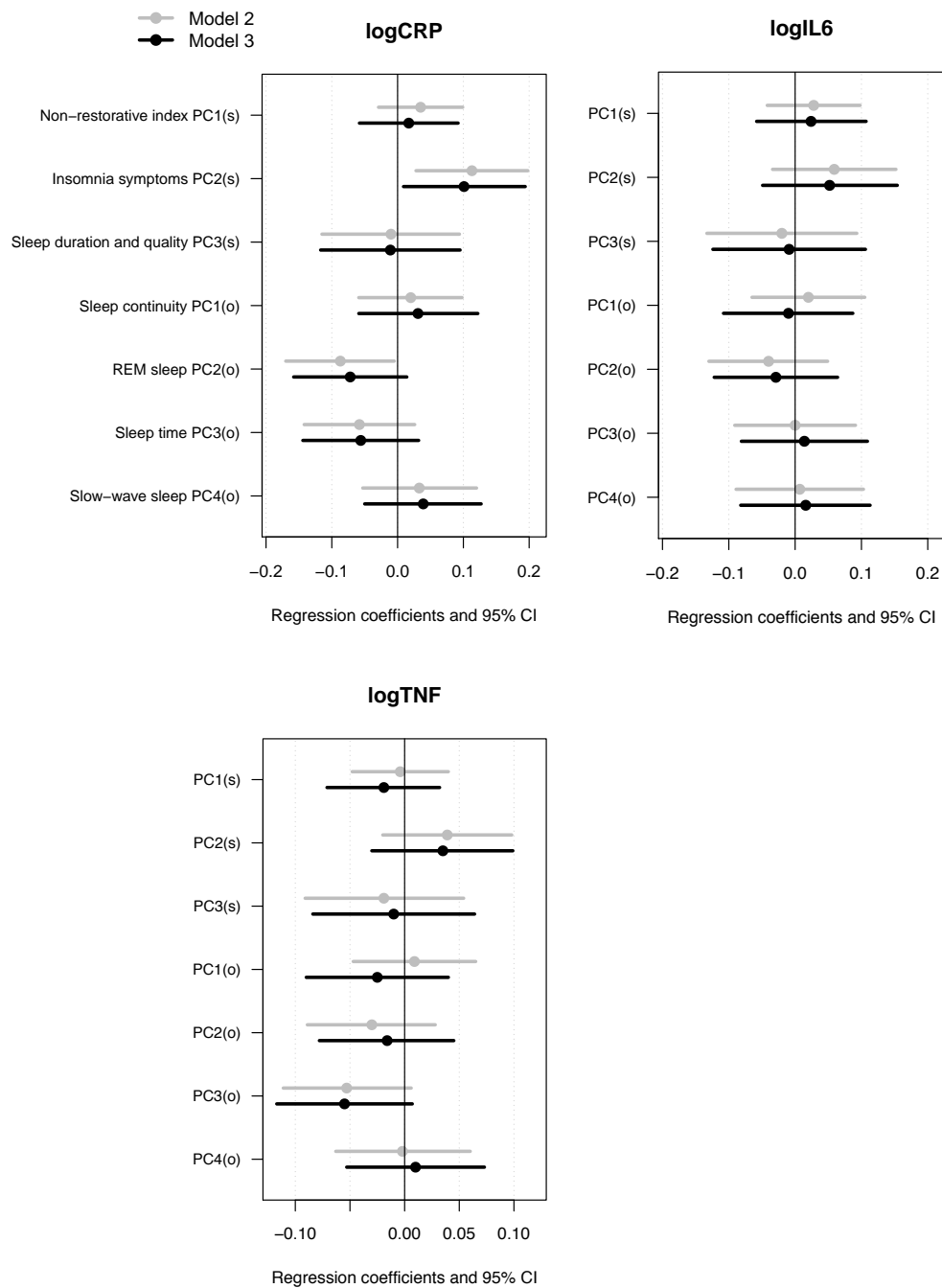
The distribution of the inflammatory markers investigated (CRP, IL-6 and  $\text{TNF}\alpha$ ) is shown in figure 6.5. As their distributions were positively skewed, we log transformed the three outcomes before conducting the analyses.



**Figure 6.5:** Distribution of C-Reactive Protein (a), Interleukin-6 (b) and Tumor Necrosis Factor  $\alpha$  (c) in 319 women included in the study.

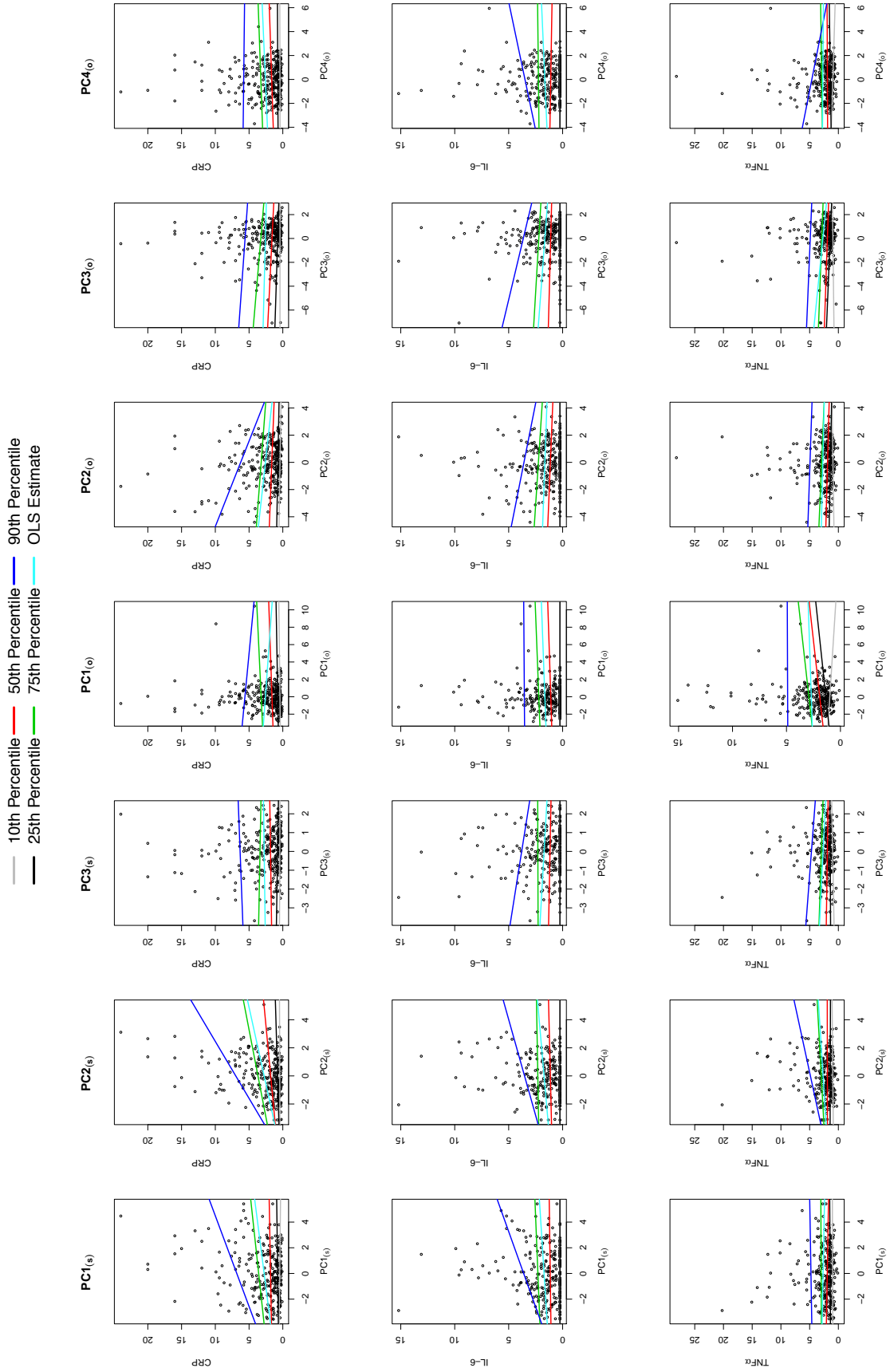
With PCA we identified three principal components for sleep variables measured through questionnaire, which together explained 66% of the original variance. Based on the correlation with original variables, the first component, PC1(s), was named "non-restorative index"; the second, PC2(s), "insomnia symptoms"; the latter PC3(s) "Sleep duration and quality". For PSG sleep variables, four principal components were retrieved: PC1(o), PC2(o), PC3(o), and PC4(o). They were named "Sleep Continuity", "REM sleep", "Sleep time", and "Slow-wave sleep", respectively.

Figure 6.6 shows the results from adjusted linear regression models with PC as exposures of interest (further details can be found in Table 2 in Manuscript IV). We found a significant association between insomnia symptoms and higher logCRP levels, but insomnia was not related to the remaining inflammatory markers. Additionally, only the REM component seemed to be negatively related with logCRP levels. When further analyzing these results, we found that the insomnia dimensions most associated with logCRP were having difficulties in maintaining sleep ( $\beta=0.33$ ; 95% CI=0.03; 0.62) and early-morning awakenings ( $\beta=0.55$ ; 95% CI=0.23; 0.88); with a subsequent 39% and 73% increase in CRP, respectively, in individuals with these complains. No association was observed with problems falling asleep ( $\beta=-0.35$ ; 95% CI=-0.77; 0.07).



**Figure 6.6:** Regression coefficients and 95% confidence intervals of the association between each principal component with the outcomes C-Reactive Protein, Interleukin-6 and Tumor Necrosis Factor  $\alpha$  (all log transformed). Model 2 is adjusted for age, BMI, physical activity, smoking status, and alcohol consumption; Model 3 is further adjusted for subjective health, apnea-hypopnea index and medication use.

As for quantile regression, it seems that the 90<sup>th</sup> percentile of CRP is influenced by insomnia symptoms, non-restorative index and the REM component (see Figure 6.7). Therefore, these variables play a role in the discrimination between high and low values of CRP. Contrarily, no statistically significant association was found with IL-6 and TNF $\alpha$ .



**Figure 6.7:** Associations between principal components and inflammatory markers, results from quantile regression analysis models adjusted for age and BMI.

# Chapter 7

## Discussion

### 7.1 Interpretation of Findings

#### 7.1.1 Obesity

Study II is one of the first studies investigating the role of obesity on the occurrence of a wide range of infections. The study confirmed a positive association between a BMI > 30 and an increased incidence of infections requiring health care contact; findings are strengthened by the fact that the association is more evident in infections requiring overnight hospitalization, a proxy of severity. Also, a dose response effect was observed in skin infections and urinary tract infections, in which not only obese, but also overweight individuals showed a higher risk compared to normal weight subjects.

The fact that the influence of obesity on the immune system might be different in men and women is interesting. Obese women are more susceptible than men, as we found an increased incidence of skin infections, gastrointestinal tract infections, urinary tract infections and sepsis; whereas only skin infections were more frequent among obese men compared to normal weight men.

The association with skin infections is well established, as it has been reported many times in previous studies (Huttunen and Syrjänen, 2013; Harpsøe et al., 2016), and it might be explained by the increased presence of skin folds in obese individuals. Findings on sepsis require further investigation. Our results extend the association found in morbidly obese individuals by Wang et al. (Wang et al., 2013), however they contrast with a previous study in which no association was found between obese women and sepsis (Kaspersen et al., 2015). The lack of association found in previous research might be due to an insufficient power of the study to detect the association due to the rare nature of this outcome.

Research on gastrointestinal infections is still limited. Our study provided the first evidence that obese women are at an increased risk of gastrointestinal tract infections, whereas this association is less clear in men. A linear trend, with a higher incidence in higher BMI values, was found in both genders, but the association was not significant in obese men. However, two recent cohort studies reported no association between obesity and gastrointestinal tract infections (Kaspersen et al., 2015; Harpsøe et al., 2016). This suggests that

further studies are warranted to confirm our findings.

Overweight and obese women showed an increased incidence of urinary tract infections compared to normal weight women, whereas in obese men, no association was observed. These findings build up a body of evidence that the association might be present only in women, as previously found in a recent study (Kaspersen et al., 2015). Some studies however contrast to our findings, in that they did not find increased incidence in women (Harpsøe et al., 2016), or found an association with urinary tract infections in both genders, but only with BMI over 50 (Saliba et al., 2013). However, no one in our study had BMI over this threshold.

As for respiratory infections, it is too early to draw any conclusions about the role of BMI. Previous studies which found a positive association between obesity and respiratory infections did not distinguish between lower or upper respiratory tract (Campitelli et al., 2014; Kaspersen et al., 2015). We found a decreased risk of lower respiratory tract infections in overweight women and a subsequent increased risk, although non-significant, in obese women. This J-shaped relationship (U-shaped if also considering underweight women) has been reported previously (Harpsøe et al., 2016). An increased risk of upper respiratory tract infections was found in Danish women with obesity (Harpsøe et al., 2016), whereas we did not find evidence in support of this association. Contrarily, upper respiratory tract infections showed a positive linear trend with BMI in men.

It is encouraging, however, that by increasing physical activity it might be possible to reduce the detrimental influence of obesity on infection risk. Our findings suggest an interaction between obesity and physical activity in such a way that only obese individuals who are physically inactive are at increased risk. In addition, our results help to discriminate across different body compositions. Only obese individuals with a waist circumference above the median presented an increased risk of infections. Although further investigations are needed, our findings offer a first line of evidence in favor of preventive measures that can be taken.

### 7.1.2 Physical Activity

In Study I we did not find evidence for an association between total physical activity and the incidence of URTI. This lack of association was also observed in a large cohort study (Hemilä et al., 2003) and in a Cochrane Systematic Review of randomized trials having a specific exercise programme as the main intervention (Grande et al., 2015). Our results, however, are in disagreement with other observational studies where physical activity was found to reduce the risk of URTI (Matthews et al., 2002; Fondell et al., 2011).

If the lack of association is the truth, or if it is due to limitations of our study which preclude the rejection of the null hypothesis, is hard to tell. It is also difficult to compare results from previous studies as they are heterogeneous in physical activity definition, ascertainment of the outcome, models used for the analyses and populations included. To cite a few examples, one study included participants over 65 years of age (Kostka et al.,

2000), one included only activities of 3 METs or more (Matthews et al., 2002), one assessed physical activity with only two categorical questions (Nieman et al., 2011), and one used cut-offs very different from ours to categorize total physical activity (<45, 45 to <55, and  $\geq 55$  METs h/d), in such a way that our reference group would be included in their lowest group (Fondell et al., 2011).

If we translate the WHO recommendations on weekly physical activity, they would correspond to an average of 750 METs minutes per week (5 times \* 30 minutes \* 5 METs), or 12.5 METs hours per week. This however, is only referring to leisure time physical activity. As there are no guidelines on how to categorize 24-hours physical activity measured as METs h/d, we subdivided it into tertiles. As a consequence, the cut-offs used are population specific. In our study, a high physical activity level ( $>43.9$  METs h/d) might correspond to an individual with a predominantly sedentary job, who goes running one hour every day and is moderately active the rest of the day, or to a construction worker who does not do any extra physical activity apart from the efforts required from his job. The difference in terms of METs h/d between the lowest and highest category is only 6 METs h/d. It is likely that this difference is too small to discriminate individuals getting a cold or not.

### 7.1.3 Sleep

Our results do not support the hypothesis of an association between sleep habits (neither sleep duration, nor sleep quality) with the occurrence or the number of URTI. In an experimental setting it has been shown that shorter sleep duration is associated with an increased risk of common cold (Cohen et al., 2009; Prather et al., 2015a). Only two studies, that we are aware of, investigated this association on a large scale. One, conducted on a cohort of female nurses, found that both short and long sleep durations are associated with an increased risk of pneumonia; the second, with a cross-sectional design, found that short sleep—but not long sleep—is associated with a greater likelihood of infections.

Our study included individuals younger than 65 years that were representative of the workforce of a city in central Sweden. It is plausible that workers are healthier than the general population and, as the immune response declines with age, it might be that sleep influences the immune system differently at different ages. Therefore, if studies include different age ranges they might find results which are not directly comparable.

More recently, long sleepers started to be in the spotlight in response of the increasing number of studies showing an association between long sleep duration and worse health outcomes (Cappuccio et al., 2010; Liu et al., 2017; Jike et al., 2018). The role of long sleep on infection risk is still under discussion as only limited research is available. In our study we did not find any evidence of an increased incidence of URTI in individuals in the higher category of sleep duration. However, we are aware of the fact that we considered individuals who reported to sleep 8 hours or more as long sleepers, and 8 hours is commonly seen as the recommended sleep duration.

In Study IV, women with insomnia symptoms showed higher levels of CRP, however no

association was found with the other inflammatory markers, IL-6 and TNF $\alpha$ . IL-6 and TNF $\alpha$  follow a circadian pattern (Vgontzas et al., 2005), in contrast to CRP levels which are quite stable over 24-hours. This could partially explain the lack of association. Inflammatory markers were measured only once through fasting blood samples taken in the morning after PSG. In order to draw more solid conclusions, studies with repeated measurements over 24 hours are desirable.

Our findings suggest that the inability to remain asleep and waking in the early morning hours are the two dimensions which lead the association between insomnia and CRP, of least importance is having sleep onset problems. The stronger association with the former symptoms might be explained by continued activation of the immune system in individuals who wake up frequently during the night. Results from previous studies are largely inconsistent. Some did not find any association between sleep disturbances and CRP in women (Liukkonen et al., 2007; Laugsand et al., 2012; Jackowska et al., 2013; Prather et al., 2015b), whereas others reported higher levels of CRP and IL-6 in women with poor sleep (Suarez, 2008; Prather et al., 2013). A meta-analysis of 72 studies (Irwin et al., 2016) showed that sleep disturbances were associated with higher levels of CRP and IL-6, but not TNF $\alpha$ . Our results are partially in line with these findings.

From quantile regression, it emerged that the largest differences were observed in the upper tail of the CRP distribution. In particular the 90<sup>th</sup> percentile of CRP is shifted upward or downward according to insomnia symptoms, non-restorative index and the REM component. As the relationship between sleep and immune systems is bidirectional, and the study has a cross-sectional design, it is also plausible that non-restorative sleep is a consequence of high CRP levels and therefore the interpretation needs cautiousness. The negative association between REM sleep and CRP levels is a novel finding. However, it has been shown that the sleep architecture in individuals with insomnia is different compared to normal sleepers in a way that the formers have a reduction in REM sleep; also, a shortened duration of REM sleep was found during the course of an infection and after administration of IL-6 (Irwin, 2019). Before inferring on the protective effect of REM sleep in regards to inflammation, experimental studies of REM deprivation and inflammation are needed to assess the directionality of the association.

A final note is to be made about the results from objective and subjective sleep measurements. Sleep duration was not associated with inflammatory markers, neither self-reported nor from PSG. This is an indication that individual judgment about sleep might be more important than sleep duration itself. The concordance between self-reported sleep and PSG, however, was poor as we did not find any association between sleep fragmentation retrieved from PSG and inflammatory markers. A plausible explanation is that the two distinct instruments are measuring different sleep characteristics, and therefore the concordance of results is not expected. Self-reported sleep usually refers to the previous 3 months and measures habitual sleep, PSG refers to that specific night which could deviate from the norm because of the “first night effect”, even if it has been shown that in-home PSG are less affected than laboratory PSG (Edinger et al., 1997).



Shorter sleep duration has been found to be related with higher IL-6, but only when objectively measured (Irwin et al., 2016). No association between self-reported short sleep duration and CRP or IL-6 was observed in a large meta-analysis, as well as in a recent study not included in the review (Irwin et al., 2016; Nowakowski et al., 2018), which goes in the same direction as our null results. Also, long sleep duration (>8 hours) was found to be associated with higher CRP and IL-6 (Irwin et al., 2016). In our study, however, only 5% of women reported to sleep more than 8 hours (same percentage from PSG), which preclude the possibility to examine the relation between long sleep and markers of inflammation.

#### 7.1.4 Stress

The association between job stress and infection occurrence that emerged from Study III is different compared to what we hypothesized initially. Our findings suggest that employees with high job demands have a higher risk of infections, especially upper respiratory tract infections and urinary tract infections, than those with less demanding jobs. However, we did not find evidence for a decreased incidence of infections in individuals with high job control. As for job strain, active jobs were associated with an increased incidence of infections compared to low strain jobs. No difference was observed in high strain jobs compared to low strain jobs.

So far, studies investigating the association between job related stress and infection occurrence mainly focused on common cold. Specifically, a positive association was found between high job demands, insufficient job control and increased risk of common cold (Mohren et al., 2001; Park et al., 2011). Our study partially confirm these findings, as we found that upper respiratory infections were more frequently occurring in employees with high demands, but the decision latitude did not play any role in the association.

Among the other infections investigated, stressed employees had a higher incidence of urinary tract infections, but no difference was observed for lower respiratory tract infections, skin infections or gastrointestinal tract infections. This is the first study supporting an association between job demands and infections of the urinary tract. It is plausible that individuals that cannot cope with the high amount of demands during their work day, feel that the time at their disposal is not enough to carry out their assignments. They are therefore more likely to hold urine for longer length of time, which itself increases the risk of infections. In addition, there might be common neuropharmacological pathways between stress and urological conditions, making the link more plausible (Sanford and Rodriguez, 2017). Despite the strength of the association found in Study III, future studies are needed to consolidate our finding.

## 7.2 Methodological Considerations

*“All models are wrong, but some are useful”*

—George E. P. Box

This is a famous quote often attributed to the British statistician George E. P. Box. More broadly, it can be applicable to epidemiological studies if changed to “All epidemiological studies are affected by errors, but some are useful”; and I would definitely say that more than *some* are useful, if not even essential. Errors can be either random or systematic, and researchers always do their best to minimize both sources of errors. Systematic errors affect validity, whereas random errors affect precision of study results. These two concepts are discussed in the subsequent paragraphs.

### 7.2.1 Validity

Validity is the feature which allows to draw inferences from study results. When validity is not an issue, it is possible to draw appropriate conclusions about the research question in individuals not included in the study sample. If results are *internally* valid, they can be extended to the source population (population from which study subjects are drawn); if results are *externally* valid, one can generalize study findings to people outside the source population.

Cohort studies, due to their observational nature, might be affected by different types of bias which can threaten the internal validity of the results. Bias is usually defined as a systematic error which leads to wrong conclusions about the true association between the exposure and the outcome. Systematic bias can be classified into three broad categories: selection bias, information bias and confounding (Rothman et al., 2008).

#### Selection bias

When study participants are not representative of those eligible to participate, selection bias might occur. In particular, the bias arises when the association between exposure and outcome among those who participated is different compared to what would have been in all eligible individuals (Rothman et al., 2008). Selection bias might follow factors that influence study participation, or it can result from differential drop-out rate of subjects during follow-up. At analysis stage it can occur if analyses are restricted on complete data only and missing is not at random. In all our studies, selection bias was a concern.

In Study I a random sample of 14,008 individuals aged 25-63 years and employed in Eskilstuna was provided by Statistics Sweden to be contacted for participating in the SWEDE-I cohort. In a previous study with similar design and commitment, an over-representation of women was observed, together with an under-representation of the age group 18-39 (Bexelius et al., 2010). For this reason, the sampling scheme used in Study I was age- and sex- stratified, with an overrepresentation of men and individuals aged 25-44 years.

Given the low participation rate observed in the previous study, 14,008 individuals were invited to achieve a planned sample size of 2,200. The reason why such a low percentage of invited individuals enrolled in the study is probably due to the heavy commitment required (Bexelius et al., 2010).

Of positive note is the fact that, having discouraged less motivated individuals to participate if they felt the efforts required by the study were too much, the drop-out rate in our study was only 3%. The missingness pattern of questionnaires' answers was somehow related to the reporting of the outcome, so that those who reported fewer URTI events had more missing answers in the questionnaires. In Study I we took this into consideration. To avoid further selection of study participants we estimated missing answers through multiple imputation before conducting the statistical analyses.

Despite all these measures that were taken to contain selection bias we cannot firmly say that the study sample is representative of the workforce of Eskilstuna. When compared to the working age Swedish population in 2011, the proportion of men and individuals in the age stratum 25–44 in the study was lower despite our efforts to recruit this group. In addition, it is important to remember that we cannot generalize our results to the whole Swedish population because of the healthy worker effect, defined as the tendency of employed individuals to be healthier than the general population.

In Studies II and III, participants are self-selected and, given the theme of the National March being related to cancer and lifestyle, they might be more health-conscious and more interested in supporting cancer research than the general population. Among those who participated in the fundraising event (whose total number could not be assessed), those who filled in the questionnaire might have additional characteristics which distinguish them from the other individuals present at the event who did not complete the questionnaire. For example, 82% of respondents, answered they walked the march which might translate in an overrepresentation of physically active individuals; and, compared to the general population, current smokers were underrepresented in the SNMC (around 10% , versus 21% of Swedish population aged 16–84 (SCB, 2018)). As a consequence, data on exposures might not reflect the true proportion in the general Swedish population.

It is not unintentional non-representativeness per se, however, which arises alertness; greater bias in the association of interest is present when the probability of self-selection in the study is associated with the outcome (Pizzi et al., 2011). Due to the prospective nature of the studies, we have no reason to believe that hospitalizations for infections occurring many months or years after enrollment might have influenced the decision to take part in the study. To conclude, the potential volunteer bias in Studies II and III has to be balanced against the high motivation of participants in answering to the questionnaire and to the essentially complete follow-up guaranteed through registers record-linkage.

In Study IV, an initial sample of 10,000 women was randomly selected from the population register of Uppsala, a Swedish city of around 160,000 inhabitants. All invitee received a postal questionnaire, and, after a series of reminders for non-respondents, the participation rate was 71.6% with a total of 7,051 women returning the questionnaire. The proportion

of women participating in the study is high, and therefore we were not concerned about representativeness at this stage.

In the second phase, 400 non-pregnant women aged 20-70 were randomly selected to undergo a home-based overnight PSG. To guarantee a sample with enough variation in the index measuring sleep apnea severity, snorers were oversampled. In this second phase, we could use PSG data from 80% of the women selected for PSG. This further selection adds to the decision in taking part into the study. However, when we compared self-reported sleep in women with complete PSG measurements ( $n=319$ ) and the whole sample ( $n=400$ ), the estimates did not change. Also, as the outcome derives from blood samples taken in the morning after PSG on all 400 women, it is not likely that the completeness of PSG measurements was related to the outcome anyhow.

### Information bias

Information bias refers to misclassification or measurement errors of the exposure and/or outcome (Rothman et al., 2008). In cohort studies, where most of the information available comes from self-reported questionnaires, is not rare to obtain inaccurate and incomplete information. In general, both the outcome and the exposure can be misclassified. If misclassification is nondifferential—namely the assessment of the exposure is similar among individuals with and without the outcome under investigation, and misclassification of the outcome is equally distributed in exposed and unexposed individuals—then this type of misclassification leads to a bias towards the null (Rothman et al., 2008). However, when exposure is not dichotomous, more complex scenarios arise which cause the direction of the bias being dependent to the categories individuals are classified to. When misclassification is differential, i.e. exposure classification depends on the outcome or viceversa, the direction of the bias is unpredictable.

In Study I, concerns on information bias were related to either the exposure or the outcome, as both were self-reported. Physical activity was measured through validated instruments (Lagerros et al., 2006; Bonn et al., 2012), but still some degree of errors cannot be excluded. A previous validation study reported a greater extent of physical activity misclassification among low educated individuals (Lagerros et al., 2006), therefore differential misclassification with respect to level of education might have occurred. In addition, as no recommended cut-offs are available, we used tertiles to categorize physical activity exposure. This might have given rise to physical activity groups whose differences in METs h/d are not biologically meaningful with the consequence of introducing more bias than what might have come from measurement errors of the instruments used. On the other hand, an arbitrary choice of cut-offs might have induced bias as well.

In the same study, we also analyzed the influence of self-reported sleep duration on the occurrence of URTI. It has been shown that there is a tendency to overestimate the amount of time spent sleeping when using self-reported measures, compared to objective sleep duration from PSG (Lauderdale et al., 2008). In particular, the overestimation is

more pronounced in short sleepers. In our study, only 1% of individuals reported a sleep duration greater than 8 hours, therefore concerns about overestimation were limited. One might question the validity of one single measurement for habitual sleep duration, however, previous studies showed a good correlation between one single question and average sleep duration obtained from 7-day diaries (Patel et al., 2004).

The greatest threat to internal validity in Study I might however be underreporting, and therefore misclassification, of the outcome. Whereas there is virtually no false positive, the total number of infections reported might be a substantial underestimate of the true number. In a study validating a similar system for reporting infectious diseases (Merk et al., 2013) it was estimated that the extent of underreporting was in the order of 50% and underreporting was more frequently observed in men, old people, and individuals with low education. In Study I we found that the number of infections correlated negatively with the number of missing answers in the exposure questionnaires. However, we do not believe there is a link between underreporting and physical activity levels or sleep habits and therefore misclassification can be considered nondifferential. Of positive note is that underreporting has been found to be constant within and between influenza seasons (Merk et al., 2013). In our study, when trends in weekly proportions of positive nasal swabs were compared with trends in the routine healthcare in Stockholm county, no substantial difference was observed (Plymoth et al., 2015).

In Study II we mainly measured obesity with self-reported BMI. It is well known from the literature that individuals tend to overestimate height and underestimate weight, which leads to an underestimation of BMI (Niedhammer et al., 2000; Gorber et al., 2007). This behavior is more evident in those with a higher BMI (Spencer et al., 2002). Because of the prospective nature of the study, differential misclassification of the exposure with respect of the outcome (infections) is unlikely. The long follow-up might raise concerns on the single measurement of BMI available in our study. It has been shown, however, that over a five-year period, the majority of individuals remain within their BMI classification (Hopman et al., 2007). We should take into account that, over longer periods, changes are more likely to occur, and they can be of either direction. It has been reported that BMI category is more likely to increase compared to decrease over time (Hopman et al., 2007), however older age groups and individuals in the higher classes of obesity can be the exceptions.

Finally, BMI does not distinguish among different body compositions. As a result, we might have classified some individuals with a low proportion of adiposity and high muscle mass as overweight, and similarly, we might have classified some individuals with a normal BMI as normal weight individuals, but missed the abdominal adiposity. In response to this, we repeated the analyses using waist circumference as the main exposure of interest.

The outcome in Study II derives from National Patient Registers. Despite the high validity of these registers (Ludvigsson et al., 2011), minor infections—seen mainly in primary care—are not covered in the registers. In addition, the outpatient register was established in 2001, therefore infections treated in outpatient care prior 2001 are not included in the study. The differential mechanism of the outcome classification with respect of obesity con-

dition does not seem to be a major issue, however, we cannot exclude that individuals with obesity, due to their condition, are more likely to be on alert and to visit an hospital when infections occur.

In study III the same reasoning applies for the outcome. Even if underestimates of the total number of infections might have occurred due to the lack of data from primary care, this is independent from the exposure to job stress. In addition, a classification of the outcome differentiating between infectious agents might have given further insight about the association.

The exposure, job strain, was measured through a Swedish version of the Demand-Control Questionnaire, which has been validated previously. The questionnaire presents distinct dimensions of demands and latitude, and the internal consistency of subscales is satisfactory (Sanne et al., 2005). As the “conflicting demands” item often has shown low loadings and has been considered a problematic item (Pelfrene et al., 2001; Sanne et al., 2005), we did not include it in the job demand score. Despite that misclassification is possible when classifying individuals using the quadrant approach, this is nondifferential with regards of the hospitalizations for infections. Having assessed job stress only once at baseline might however be a limitation. Individuals are likely to change jobs and therefore change their level of stress as a result. We hypothesized that individuals exposed to high levels of job strain are more likely to turn into jobs requiring less strain than the opposite, and that this should happen independently of the outcome with subsequent underestimation of the association of interest. As we censored participants at their retirement, we could limit the misclassification bias. Individuals who retired during follow-up, and therefore were not exposed to occupational stress anymore, no longer contributed person-time to the study after retirement.

In Study IV misclassification bias was of less concern. Sleep was assessed both subjectively and objectively and the outcome was derived from laboratory testing. Despite that, the presence of a single measurement of inflammatory markers might undermine the validity of our findings for IL-6 and TNF $\alpha$ , as they are known to vary over 24 hours. Additionally, the principal components we used, might have hidden the association with single items and markers of inflammation.

## Confounding

Confounding is one of the major threats to validity in epidemiological studies. By definition, a confounder is a factor which is associated with the outcome and it is not equally distributed across exposure categories (Rothman et al., 2008). Also, a variable, to be defined as being a confounder, needs not to be on the causal pathway between the exposure and the disease, namely it should be associated with the exposure, but must not be an effect of it. In the presence of confounding, we might observe a spurious association between the exposure and the outcome even when no causal effect exists. More generally, a confounder mixes up the observed effects. It can cause underestimates, overestimates, and when strong enough it

might also reverse the direction of the true effect of the exposure on the outcome. Methods to deal with confounding are available at all stages, design and analyses, but the most commonly used are the multivariate regression models.

In Study I the associations between our exposures of interest, sleep and physical activity, and the outcome were adjusted for age, sex, presence of children  $\leq 6$  years old in the household, smoking, travel model to work, education, BMI, number of close contacts per day, treatment for allergy, asthma or lung cancer, and treatment for immunodeficiency or transplantation. The selection of potential confounders was done based on subject matter knowledge. In a crude analysis, long sleep and bad quality sleep were factors associated with a greater number of URTI, but the association went away when adjusting for other variables, indicating that the association might have been spurious.

In Study II, based on a priori expert knowledge (Dobner and Kaser, 2018), we adjusted the models for age, smoking status, alcohol consumption, education, disposable income, physical activity, sleep duration, Charlson's comorbidity index and hormone treatment (women only). Despite that we collected data from the food frequency questionnaire on dietary habits, and diet might be associated with both obesity and infections (Farhadi et al., 2018), we cannot exclude that dietary habits are influenced by BMI. Additionally, due to the high presence of missing values on the food items frequencies and the uncertainty around the best measure to use to control for diet (e.g. nutrients, macronutrients, dietary patterns), we decided not to adjust our estimates for it.

In Study III, as described in Chapter 5, we used a DAG to select the minimal set of variables to adjust our models for (Supplementary Figure 1 in Manuscript III). The DAG was drawn based on a priori assumptions on the relationship between exposure, intermediate variables, confounders and the outcome. Manuscript III presents the estimates resulting from three different models: one age and sex adjusted; the second with the addition of variables suggested by the DAG; the third model also included the potential intermediate variables. Even if we used this emerging method to select covariates, the DAG strongly relies on the researcher's assumptions and subject matter knowledge. Assuming the DAG is true, the results are valid. In real life, however, not only one DAG is true. Different researchers might have different hypotheses and come up with different DAGs, especially in terms of the directionality of the arrows when bidirectional relationships are known in the literature. In our study, all exposure information was collected at baseline and therefore the time order between variables cannot be established. Is it job stress which affects BMI, sleep duration, smoking status and alcohol consumption, or is the other way around? We cannot give a firm answer to this question, but we hypothesized that the first scenario was true.

Study IV is restricted to women; therefore, it automatically takes into account the potential confounding effect of sex by restriction of the study population. Three different models were fitted. Model 1 adjusted for age and BMI (not presented in this thesis); Model 2 additionally adjusted for physical activity, smoking status, alcohol consumption; Model 3 also included subjective health, apnea-hypopnea index and medication use. The reason why we adopted this approach is that in this way we could examine the effect of the exposure

on the outcome step by step, avoiding over-adjusting our models (Schisterman et al., 2009).

Although the availability of extensive data collected from questionnaires allowed us to control our analyses for a wide range of factors, residual confounding cannot be excluded in any of the studies included in this thesis. This is mainly because of the observational nature of our studies. Residual confounding can be due to unmeasured potential confounders, or to the categorization of continuous factors which lost precision when entered into the model as categorical. Nonetheless, we believe that we did our best to limit the extent of residual confounding, and that our estimates would not change much even after extensive control for further variables.

### 7.2.2 Precision

Even when all systematic errors are eliminated, results from epidemiological studies are not error-free. Random errors are due to unexplained variability in the data and affect precision of estimates. When a measure of association is reported, it is always followed by a confidence interval with a certain level of confidence (95% most of the times). Precision is influenced by sample size, the greater the sample the narrower the confidence interval; but it also depends on how common the exposure and/or the outcome are. By design the random error was higher in Study IV because of its smaller sample size compared to other studies, especially studies II and III including several thousand participants. In these latter studies, the precision was reduced when conducting analyses by subgroup of infections.

## 7.3 Ethical Considerations

In the last decade, register-based research has rapidly grown due to the increasing availability and improved quality of population-based registers. The valuable amount of information collected, and the remarkable cost and time savings that administrative data implies, paved the way to this appealing area of epidemiology.

In Studies II and III, we used a combination of data derived both from questionnaires filled in by SNMC participants and data from a series of nationwide and essentially complete registers. Registers were linked to self-reported data through the PIN provided by each participant at baseline. Participation into this cohort was voluntary based and subjects were informed on what were the aims of the research. The Regional Ethics Review Board in Stockholm approved the study design (Dnr 1997-205 and Dnr 2017/796-31) and informed consent was obtained from all participants included. PIN was provided by each participant, and it was a prerequisite to link data with the registers. However, when we received data from Statistics Sweden and from the National Board of Health and Welfare, data were de-identified and each individual was assigned a newly created identifier to allow the linkage across registers.

In Studies I and IV, laboratory data were also available. In Study I nasal swabs were



sent by participants during a 9-month follow-up period, whereas in Study IV blood samples were taken. The SWEDE-I study has been approved by the Regional Ethics Review Board in Stockholm (Dnr 2011/360-31/2), and the polysomnography study within the Sleep and HHealth in Women cohort was approved by the Ethics Committee at the Medical Faculty at Uppsala University (Dnr 01-238). When dealing with biological samples, further ethical considerations ought to be mentioned (Budimir et al., 2011), among which the informed consent, the storage of the samples and their re-use is important. In the SWEDE-I study, nasal swabs were stored at -70°C until tested for 14 different viruses, and the remaining material was stored in a biobank for possible future supplementary analyses. In the SHE study, blood samples were collected the morning after polysomnography and are stored in a biobank for potential future analyses.

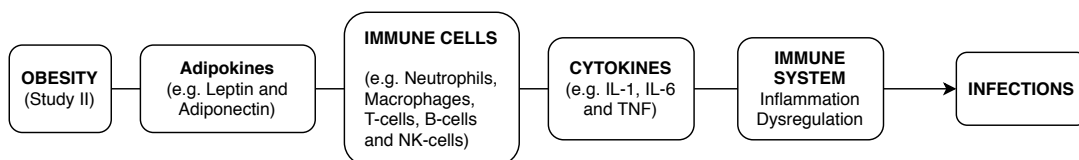
Informed consent must be obtained by participants once they are informed and have understood the purpose of the research and agreed to participate. If a second consent needs to be obtained when biological samples are re-used is object of debate. The majority of authors who contributed to studies addressing biobank ethical aspects, support the broad consent (Hansson, 2009), in which broad, but not detailed information about the research is included. In our studies, however, there is no re-use of the samples.

# Chapter 8

## Conclusions

The results presented in this thesis contribute to the body of scientific evidence regarding the association between lifestyle and infections. The initial hypotheses have been reprinted here, using dashed segments when we did not have enough evidence to reject the null hypotheses, and solid lines when the associations were confirmed. As we investigated the overall picture and not every single step of the underlying pathway, we used a unique arrow (going through the hypothesized intermediate steps) to represent the conclusions of each study. See Figures 8.1-8.4.

### 8.1 Obesity and Infections



**Figure 8.1:** Hypotheses and conclusions from Study I

There is evidence of an association between obesity and infections (Study II). The increased incidence of infections in individuals with obesity, however, is different in men and women and it varies according to type of infection. More specifically, there is an increased incidence of skin infections in both men and women with obesity, and an increased incidence of gastrointestinal tract infections, urinary tract infections and sepsis in women only.

### 8.2 Physical Activity and Infections

Contrarily to our hypotheses (Figure 8.2), total physical activity does not appear to explain differences in rate of upper respiratory tract infections (Study I).

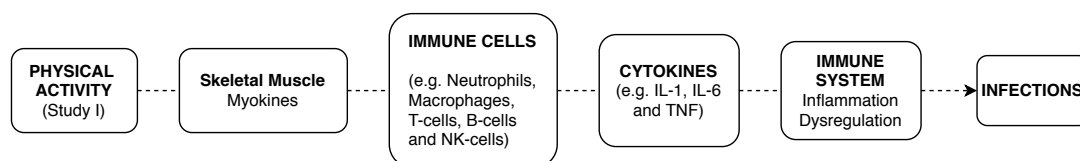


Figure 8.2: Hypotheses and conclusions from Study II

### 8.3 Sleep and Infections

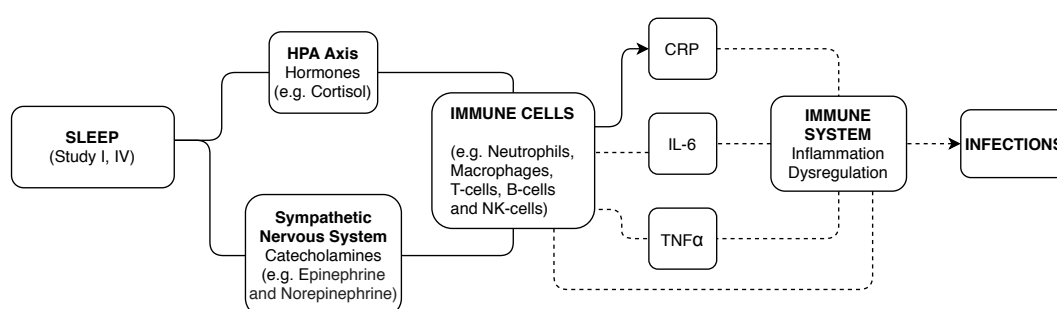


Figure 8.3: Hypotheses and conclusions from Study III

Sleep duration and sleep quality do not seem to be related with the occurrence of upper respiratory tract infections (Study I). Insomnia symptoms and REM sleep might have a role in determining the levels of CRP (insomnia increases CRP levels, whereas REM sleep diminishes them), however, their role with the other inflammatory markers, IL-6 and TNF $\alpha$ , is not clear (Study IV).

### 8.4 Stress and Infections

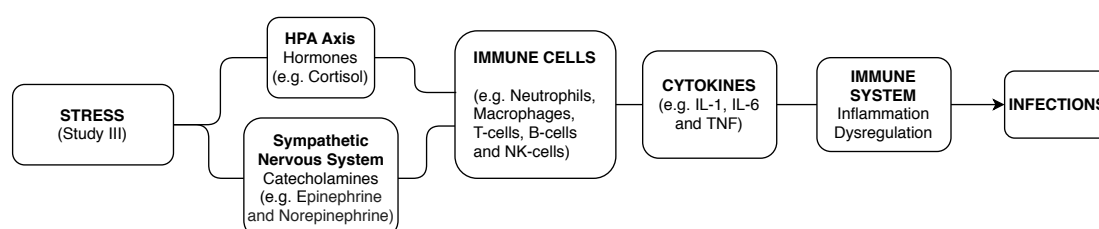


Figure 8.4: Hypotheses and conclusions from Study IV

High job demands are associated with an increased incidence of infections—particularly upper respiratory tract infections and urinary tract infections—but job control does not seem to play a role in mitigating the association.

## Chapter 9

### Future Research

Future research is needed to continue investigating the role of lifestyle on the occurrence and spread of infections. Particularly:

- The rapid development of technology paved the way to the use of smartphone measurements of physical activity in epidemiological studies. Although the validity of these instruments is still under discussion, they might in the future be able to provide objective measurements of total physical activity on a large scale and at low costs. Therefore, future studies in this setting might help in answering still-open questions on the role of physical activity in respiratory infection occurrence. Additionally, smartphones can be a way to collect real time data on infections too. Through apposite applications, for example, it would be possible to send information as soon as one feels sick.
- Further studies, differentiating by infectious agent, might be helpful to draw more solid conclusions about the influence of obesity and job stress on the occurrence of infections. It is therefore of interest to explore if the association with lifestyle is different in viral and bacterial infections.
- The association found between REM sleep and CRP levels has not been reported before and it raises the need for further studies investigating REM deprivation and inflammation to determine, if any, a direction of causality.
- Diet is another important lifestyle behavior and modifiable risk factor which plays a key role in the prevention of chronic diseases. It would be interesting to investigate which are the dietary patterns associated with the lowest incidence of infections.
- In the light of the new pandemic caused by Covid-19 it might be interesting to identify with more certainty who are the individuals more susceptible to the virus and who are the individuals with worst outcomes once infected. In this regard, a multi-source score including comorbidities, demographics and lifestyle can be developed.

# Chapter 10

## Acknowledgements

There are many people I would like to thank for their guidance, inspiration, encouragement and support throughout this PhD journey. In particular, I would like to express my sincere gratitude to the following:

**Ylva Trolle Lagerros**, my main supervisor, you have believed in me since the beginning. Thank you for giving me the possibility of doing this PhD. I am truly grateful and I have really enjoyed being a PhD student under your supervision. Thank you for your contagious enthusiasm for doing research and for your infinite energy. Thank you for having taught me the importance of being a researcher, but also that family matters, and that a balance between work and life is fundamental. Thank you for your continuous support, for having let me part of your family, for our conversations about research late in the night, for your delicious bread and for having introduced me to the authentic Swedish food. I will always treasure all what I have learnt during these years as a PhD student.

I feel honored for having had the opportunity to work closely to a worldwide known sleep researcher, **Torbjörn Åkerstedt**, my co-supervisor. You have introduced me to the field of sleep epidemiology and I am fascinated by your dedication to science. Thank you for sharing your vast knowledge with me, and for the time and patience you have dedicated to me. You are a brilliant scientist.

**Rino Bellocco**, my co-supervisor. All this would not have been possible without you. You saw potential in me as a researcher and teacher before I believed it was possible. Thank you for having encouraged me to look for job opportunities abroad just after my master thesis, and for having let me discover the beauty of doing research in an international and such stimulating environment. Thank you for having always been there in case I needed, both for research advices and personal life-related matters. Thank you for all the teaching opportunities you gave me. Despite my continuous refusals, I acknowledge that I learned a lot, and every time after class I felt my efforts rewarded.

**Amelie Plymoth**, my co-supervisor. Thank you for your kindness and for always finding time for me when I needed advice. Thank you for our productive lunches, for your constant

feedbacks, and for introducing me to your research group.

**Anna Thorson**, my co-supervisor. Thank you for agreeing to be part of this big team, and for your important tips on my last project.

**Bénédicte Delcoigne**, my PhD mentor. Thank you for caring about my research, my feelings and for catching up every time we randomly met at journal clubs, seminars, or conferences. Thank you for always sharing your experience and for your encouraging words.

I am grateful to all the co-authors for thoughtful comments, and for invaluable contributions to our papers. Especially, **Hans-Olov Adami**—every time I received your feedbacks I was amazed by how well you could synthesize concepts which took me entire sentences. Thank you also for all encouragements after each paper rejection. **Olof Nyrén** for letting me work on the SWEDE-I data and for your precious inputs. **Henric Winell** and **Andrea Bellavia** for everything you have taught me about statistics. **Hongwei Zhao**, for all your reasonings about statistics and SNMC, but also for your tasty Chinese food and fun moments together. **Weimin Ye**, for your epidemiological expertise. **Eva Lindberg** and **Jenny Theorell-Haglöw** for sharing the SHE data and for welcoming me in Uppsala.

**Bozenna Iliadou**, for help with extracting data from SWEDE-I database.

**Michael Fored**, for being an excellent study director. **Lillemor Melander**, for always answering so promptly to all my enquiries, and for taking care of hundreds of MedSolna's doctoral students so kindly. All the outstanding administrators at KEP, you have always been available and helpful, **Eva Willis**, **Helena Nord**, **Monica Rundgren**, and **Pernilla Appelquist**. I would like to thank also **Gunilla Nilsson Roos** and **Marie Jansson**, administrators at MEB, for their patience and for helping me sorting out so many things during these four years.

I would like to express my gratitude to **Marcello Pagano**, **Marco Bonetti**, and **Rino Bellocchio**, for the opportunity to take part in the Summer School on Modern Methods in Biostatistics and Epidemiology, for involving me in the organization of the school, and for the privilege to serve as teaching assistant for some of the courses.

I would like to thank the Biostat-2 teaching assistants over the past years, who together with me, have worked hard to make logistic regression more appealing to students by helping Rino in his class. Especially, **Alessandro**, **Jacopo**, **Marco**, **Ottavia**, **Shadi**, **Tobba**, and **Tyra**.

**Oskar** and **André**, my first two students. Helping you as co-supervisor in your degree projects has been a learning experience for me.

The **Biostat Group** at MEB, for having welcomed me at their Thursday's seminars.

**Stephanie**, for greeting me with a smile every time we met.

**Jonas Ludvigsson**, thank you for your efforts in rigorously explaining the content of Swedish Registers in your publications. Also, people in my village in Italy still remember the football match of your team against the local one, it was a success!

**Daniela**, the first friend I met in Stockholm. As soon as I landed in Sweden, where everything was new and so different from Italy, you were there waiting for me at the bus stop. Since then, you have always made me feel like at home, you have warmly welcomed me in your office, you have spoiled me with your desserts, and you have taught me so many things. It was impossible to get bored when you were around, climbing, ice-skating, and swimming were daily scheduled. Thank you.

**Alessandra**, a big thank you to you! You have always been in front-line listening to all my problems, you have always helped me analyzing all possible scenarios and finding the best solutions, no matter the topic. I will never forget our hours-long chats in front of a cup of tea. Thank you for being such a sweet, wise, and caring person.

**Essi**, I feel extremely lucky to have you as my colleague and friend. The experiences we have shared in the past four years have been precious to me. While writing these words, there are so many memories coming up in my mind! We have supported each other in all stages of our PhDs, we worked hard, we had fun, we have laughed to tears. Your calm and positivity is a treasure for anyone around you.

**Alessio, Andrea, Elisa, Laura, Marco, Marica**, the Italian crowd at MEB. I am grateful to all of you, your genuine friendship helped me forget I was away from home. Our coffee breaks, fun lunches, and interesting conversations made the work environment such a pleasant place to be.

I would like to thank also **Elena**, with whom I attended several PhD courses, and **Romina** for all her advices at the beginning of my Swedish experience.

**Jiayao**, we shared this adventure from the beginning. Thank you for all your help throughout this doctorate. I will miss our fun dinners, and taking pictures all around Stockholm.

Current and former PhD students at MEB, especially **Andreas, Elisabeth, Emilio, Gabriel**, and **Hannah**.

**Eleonora**, my PhD partner. Thank you for your valuable friendship, for our daily discussions about science and life, and for sharing all the ups and downs of my PhD journey. It would not have been the same without all your support and help.

My office-mate **Cecilia**. Not only we have shared the office, but also thoughts and laughs. Thank you for your cheerfulness and support. Thank you also to **Gianmarco** and **Marta**. Since you started working in our group, I stopped worrying about forgetting the office keys at home: there was always someone arriving in the office before me. I would also like to

thank the other colleagues at University of Milano-Bicocca, for many discussions over long lunch breaks. Particularly, **Anna**, **Annalisa**, **Federico**, **Giulia**, **Matteo F**, **Matteo M**, **Nicola** and **Samuele**.

My friends from Como for making sure I had a social life besides my PhD. **Vanessa**, we have grown up together and despite the distance we have not stopped encouraging each other. Thank you for always making me smile, you are like a sister to me. **Martina**, thank you for all your advices and our long talks on everything. **Alessio**, who—like me—has never stopped studying. Thank you for always being proud of me and for all your psychological support. Thanks also to **Arianna**, **Daniela**, **Elisa**, and **Valentina** for all these years of friendship. Finally, thanks to **Matteo**, **Piero**, **Riccardo**, and **Federica** for the great company during weekends.

Last, but not least, my Family.

A very special word of thanks goes to my parents for their unconditional love, trust and support throughout my life. You taught me the value of studying and always wanted the best for me. **Mamma**, the strongest women I have ever met. You fight a silence battle everyday, but you never stopped smiling and encouraging people around you. **Papi**, the sacrifices you made for our family are beyond words. Your dedication to work is exemplary and it is something that has always encouraged me to give it my all. My brother **Pietro**, we are very much alike. I admire you for daring to follow your dreams, and we are all proud of your efforts and achievements. I know I can always count on you, thank you.

My grandparents **Maria** and **Giuseppe**, you are always available to help me. I am grateful for all the good values you taught me, for showing that happiness is found in the little things, and for the extraordinary family you have created. My grandmother **Anastasia**, you have always been so engaged in every step of the way and always curious to know what I was doing and how I was proceeding. I miss you so much. To the rest of my big family, to all cousins, uncles and aunts for their continuous support and for always cheering for me. Thanks also to my mother-in law and father-in-law for being there for me and for their silent support.

My beloved husband, **Francesco**. You have enriched my life with your presence. Thank you for your generous love and understanding. You are my strength.

This work was supported by **Karolinska Institutet**'s funding for doctoral students (KID-funding)



# References

- Agresti, A. 2003. *Categorical data analysis*, vol. 482. John Wiley & Sons.
- Ainsworth, B. E., W. L. Haskell, A. S. Leon, J. D. Jacobs, H. J. Montoye, J. F. Sallis, and J. R. Paffenbarger. 1993. Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and science in sports and exercise* 25(1): 71–80.
- Amorim, L. D., and J. Cai. 2015. Modelling recurrent events: a tutorial for analysis in epidemiology. *International journal of epidemiology* 44(1): 324–333.
- Andersen, P K., and R. D. Gill. 1982. Cox's regression model for counting processes: a large sample study. *The annals of statistics* 1100–1120.
- Barlow, L., K. Westergren, L. Holmberg, and M. Talbäck. 2009. The completeness of the Swedish Cancer Register: a sample survey for year 1998. *Acta Oncologica (Stockholm, Sweden)* 48(1): 27–33.
- Bauman, A., B. E. Ainsworth, J. F. Sallis, M. Hagströmer, C. L. Craig, F. C. Bull, M. Pratt, K. Venugopal, J. Chau, M. Sjöström, et al. 2011. The descriptive epidemiology of sitting: a 20-country comparison using the International Physical Activity Questionnaire (IPAQ). *American journal of preventive medicine* 41(2): 228–235.
- Besedovsky, L., T. Lange, and J. Born. 2012. Sleep and immune function. *Pflügers Archiv-European Journal of Physiology* 463(1): 121–137.
- Besedovsky, L., T. Lange, and M. Haack. 2019. The sleep-immune crosstalk in health and disease. *Physiological reviews* 99(3): 1325–1380.
- Bexelius, C., H. Merk, S. Sandin, O. Nyrén, S. Kühlmann-Berenzon, A. Linde, and J.-E. Litton. 2010. Interactive voice response and web-based questionnaires for population-based infectious disease reporting. *European journal of epidemiology* 25(10): 693–702.
- Beyerlein, A. 2014. Quantile regression—opportunities and challenges from a user's perspective. *American journal of epidemiology* 180(3): 330–331.
- Bonn, S. E., Y. T. Lagerros, S. E. Christensen, E. Möller, A. Wright, A. Sjölander, and K. Bälter. 2012. Active-Q: validation of the web-based physical activity questionnaire using doubly labeled water. *Journal of medical Internet research* 14(1): e29.

- Borbély, A. A. 1982. A two process model of sleep regulation. *Human neurobiology* 1(3): 195–204.
- Boscolo, P., A. Di Donato, L. Di Giampaolo, L. Forcella, M. Reale, V. Dadorante, F. Alparone, S. Pagliaro, M. Kouri, A. Magrini, et al. 2009. Blood natural killer activity is reduced in men with occupational stress and job insecurity working in a university. *International archives of occupational and environmental health* 82(6): 787–794.
- Boscolo, P., L. Forcella, M. Reale, G. Vianale, U. Battisti, R. Bonfiglioli, M. Cortini, L. Di Giampaolo, A. Di Donato, and S. Salerno. 2012. Job strain in different types of employment affects the immune response. *Work* 41(Supplement 1): 2950–2954.
- Brooke, H. L., M. Talbäck, J. Hörnblad, L. A. Johansson, J. F. Ludvigsson, H. Druid, M. Feychting, and R. Ljung. 2017. The Swedish cause of death register. *European journal of epidemiology* 32(9): 765–773.
- Bryant, P. A., J. Trinder, and N. Curtis. 2004. Sick and tired: does sleep have a vital role in the immune system? *Nature Reviews Immunology* 4(6): 457–467.
- Budimir, D., O. Polašek, A. Marušić, I. Kolčić, T. Zemunik, V. Boraska, A. Jerončić, M. Boban, H. Campbell, and I. Rudan. 2011. Ethical aspects of human biobanks: a systematic review. *Croatian medical journal* 52(3): 262–279.
- Campitelli, M. A., L. C. Rosella, and J. C. Kwong. 2014. The association between obesity and outpatient visits for acute respiratory infections in Ontario, Canada. *International journal of obesity* 38(1): 113–119.
- Cappuccio, F. P., L. D’Elia, P. Strazzullo, and M. A. Miller. 2010. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep* 33(5): 585–592.
- Caspersen, C. J., K. E. Powell, G. M. Christenson, et al. 1985. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health rep* 100(2): 126–31.
- CDC. 2014. Centers for Disease Control and Prevention. Sleep and Sleep Disorders. Data and Statistics URL [https://www.cdc.gov/sleep/data\\_statistics.html](https://www.cdc.gov/sleep/data_statistics.html) (Accessed February 3, 2020).
- Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases* 40(5): 373–383.
- Christensen, K. L., R. C. Holman, C. A. Steiner, J. J. Sejvar, B. J. Stoll, and L. B. Schonberger. 2009. Infectious disease hospitalizations in the United States. *Clinical Infectious Diseases* 49(7): 1025–35.

- Cohen, S. 1996. Psychological stress, immunity, and upper respiratory infections. *Current Directions in Psychological Science* 5(3): 86–89.
- Cohen, S., W. J. Doyle, C. M. Alper, D. Janicki-Deverts, and R. B. Turner. 2009. Sleep habits and susceptibility to the common cold. *Archives of internal medicine* 169(1): 62–67.
- Cox, D. R. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)* 34(2): 187–220.
- Dhabhar, F. S., and B. S. McEwen. 1997. Acute stress enhances while chronic stress suppresses cell-mediated immunity in vivo: A potential role for leukocyte trafficking. *Brain, behavior, and immunity* 11(4): 286–306.
- Dobner, J., and S. Kaser. 2018. Body mass index and the risk of infection—from underweight to obesity. *Clinical Microbiology and Infection* 24(1): 24–28.
- Edinger, J. D., A. I. Fins, R. J. Sullivan Jr, G. R. Marsh, D. S. Dailey, T. V. Hope, M. Young, E. Shaw, D. Carlson, and D. Vasilas. 1997. Sleep in the laboratory and sleep at home: comparisons of older insomniacs and normal sleepers. *Sleep* 20(12): 1119–1126.
- Elliott, G. R. 1982. Stress and human health. Analysis and implications of research. *A study by the Institute of Medicine, National Academy of Sciences*.
- EU-OSHA. 2014. Calculating the cost of work-related stress and psychosocial risks. Technical report, European Agency for Safety and Health at Work. *Publications Of- fice of the European Union*. URL <https://osha.europa.eu/en/publications/calculating-cost-work-related-stress-and-psychosocial-risks>
- Eurofound. 2017. Sixth European Working Conditions Survey—Overview re- port (2017 update). *Publications Office of the European Union, Luxembourg*. URL [https://www.eurofound.europa.eu/sites/default/files/ef\\_publication/field\\_ef\\_document/ef1634en.pdf](https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef1634en.pdf)
- Falagas, M. E., and M. Kompoti. 2006. Obesity and infection. *The Lancet infectious diseases* 6(7): 438–446.
- Farhadi, S., R. S. Ovchinnikov, et al. 2018. The relationship between nutrition and infectious diseases: A review. *Biomedical and Biotechnology Research Journal (BBRJ)* 2(3): 168.
- Ferrie, J. E., M. Kumari, P. Salo, A. Singh-Manoux, and M. Kivimäki. 2011. Sleep epidemi- ology—a rapidly growing field. *International Journal of Epidemiology* 40(6): 1431–7.
- Fondell, E., Y. T. Lagerros, C. J. Sundberg, M. Lekander, O. Bälter, K. J. Rothman, and K. Bälter. 2011. Physical activity, stress, and self-reported upper respiratory tract infection. *Medicine & Science in Sports & Exercise* 43(2): 272–279.

- GBD 2016 Lower Respiratory Infections Collaborators, et al. 2018. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet. Infectious diseases* 18(11): 1191.
- GBD 2017 Risk Factor Collaborators, et al. 2018. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England)* 392(10159): 1923.
- Ghilotti, F., A. Julander, P. Gustavsson, A. Linde, O. Nyrén, and A. Plymoth. 2019. Cohort profile: Studies of Work Environment and Disease Epidemiology-Infections (SWEDE-I), a prospective cohort on employed adults in Sweden. *PloS One* 14(5).
- Glaser, R., and J. K. Kiecolt-Glaser. 2005. Stress-induced immune dysfunction: implications for health. *Nature Reviews Immunology* 5(3): 243–251.
- Gorber, S. C., M. Tremblay, D. Moher, and B. Gorber. 2007. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity reviews* 8(4): 307–326.
- Grande, A. J., J. Keogh, T. C. Hoffmann, E. M. Beller, and C. B. Del Mar. 2015. Exercise versus no exercise for the occurrence, severity and duration of acute respiratory infections. *Cochrane Database of Systematic Reviews* (6).
- Grimsmo, A., E. Hagman, E. Faikø, L. Matthiessen, and T. Njálsson. 2001. Patients, diagnoses and processes in general practice in the Nordic countries. An attempt to make data from computerised medical records available for comparable statistics. *Scandinavian Journal of Primary Health Care* 19(2): 76–82.
- Guo, Z., T. M. Gill, and H. G. Allore. 2008. Modeling repeated time-to-event health conditions with discontinuous risk intervals. *Methods of information in medicine* 47(02): 107–116.
- Guthold, R., G. A. Stevens, and F. C. Riley, Leanne M fand Bull. 2018. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. *The Lancet Global Health* 6(10): e1077–e1086.
- Hansson, M. G. 2009. Ethics and biobanks. *British journal of cancer* 100(1): 8–12.
- Harpsøe, M. C., N. M. Nielsen, N. Friis-Møller, M. Andersson, J. Wohlfahrt, A. Linneberg, E. A. Nohr, and T. Jess. 2016. Body mass index and risk of infections among women in the Danish National Birth Cohort. *American journal of epidemiology* 183(11): 1008–1017.
- Hassard, J., K. R. Teoh, G. Visockaite, P. Dewe, and T. Cox. 2018. The cost of work-related stress to society: A systematic review. *Journal of occupational health psychology* 23(1): 1.

- Hellgren, J., A. Cervin, S. Nordling, A. Bergman, and L. Cardell. 2010. Allergic rhinitis and the common cold—high cost to society. *Allergy* 65(6): 776–783.
- Hemilä, H., J. Virtamo, D. Albanes, and J. Kaprio. 2003. Physical activity and the common cold in men administered vitamin E and  $\beta$ -carotene. *Medicine & Science in Sports & Exercise* 35(11): 1815–1820.
- Hetta, J., J.-E. Broman, and L. Mallon. 1999. Evaluation of severe insomnia in the general population—Implications for the management of insomnia: Insomnia, quality of life and healthcare consumption in Sweden. *Journal of Psychopharmacology* 13(4 Suppl 1): S35–S36.
- Hill, J. O., H. R. Wyatt, and J. C. Peters. 2012. Energy balance and obesity. *Circulation* 126(1): 126–132.
- Hofstetter, H., E. Dusseldorp, A. Zeileis, and A. A. Schuller. 2016. Modeling caries experience: advantages of the use of the hurdle model. *Caries research* 50(6): 517–526.
- Hopman, W. M., C. Leroux, C. Berger, L. Joseph, S. I. Barr, J. C. Prior, M. Harrison, S. Poliquin, T. Towheed, T. Anastassiades, et al. 2007. Changes in body mass index in Canadians over a five-year period: results of a prospective, population-based study. *BMC Public Health* 7(1): 150.
- Huttunen, R., and J. Syrjänen. 2013. Obesity and the risk and outcome of infection. *International journal of obesity* 37(3): 333–340.
- Irwin, M. R. 2015. Why sleep is important for health: a psychoneuroimmunology perspective. *Annual review of psychology* 66: 143–172.
- Irwin, M. R. 2019. Sleep and inflammation: partners in sickness and in health. *Nature Reviews Immunology* 19(11): 702–715.
- Irwin, M. R., and S. W. Cole. 2011. Reciprocal regulation of the neural and innate immune systems. *Nature Reviews Immunology* 11(9): 625–632.
- Irwin, M. R., R. Olmstead, and J. E. Carroll. 2016. Sleep disturbance, sleep duration, and inflammation: a systematic review and meta-analysis of cohort studies and experimental sleep deprivation. *Biological psychiatry* 80(1): 40–52.
- Itani, O., M. Jike, N. Watanabe, and Y. Kaneita. 2017. Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression. *Sleep medicine* 32: 246–256.
- Jackowska, M., M. Kumari, and A. Steptoe. 2013. Sleep and biomarkers in the English Longitudinal Study of Ageing: associations with C-reactive protein, fibrinogen, dehydroepiandrosterone sulfate and hemoglobin. *Psychoneuroendocrinology* 38(9): 1484–1493.

- Jastreboff, A. M., C. M. Kotz, S. Kahan, A. S. Kelly, and S. B. Heymsfield. 2019. Obesity as a disease: the obesity society 2018 position statement. *Obesity* 27(1): 7–9.
- Jike, M., O. Itani, N. Watanabe, D. J. Buysse, and Y. Kaneita. 2018. Long sleep duration and health outcomes: A systematic review, meta-analysis and meta-regression. *Sleep medicine reviews* 39: 25–36.
- Karasek Jr, R. A. 1979. Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative science quarterly* 285–308.
- Karlsson, E. A., and M. A. Beck. 2010. The burden of obesity on infectious disease. *Experimental biology and medicine* 235(12): 1412–1424.
- Kasapis, C., and P. D. Thompson. 2005. The effects of physical activity on serum C-reactive protein and inflammatory markers: a systematic review. *Journal of the American College of Cardiology* 45(10): 1563–1569.
- Kaspersen, K. A., O. B. Pedersen, M. S. Petersen, H. Hjalgrim, K. Rostgaard, B. K. Møller, C. Juul-Sørensen, S. Kotzé, K. M. Dinh, L. T. Erikstrup, et al. 2015. Obesity and risk of infection: results from the Danish Blood Donor Study. *Epidemiology* 26(4): 580–589.
- Kivimäki, M., and A. Steptoe. 2018. Effects of stress on the development and progression of cardiovascular disease. *Nature Reviews Cardiology* 15(4): 215.
- Kleinbaum, D. G., and M. Klein. 2012. *Survival Analysis. A Self-Learning Text*. 3rd ed. New York, NY: Springer Verlag.
- Knight, J. A. 2012. Physical inactivity: associated diseases and disorders. *Annals of Clinical & Laboratory Science* 42(3): 320–337.
- Koenker, R., and K. F. Hallock. 2001. Quantile regression. *Journal of economic perspectives* 15(4): 143–156.
- Kom, E. L., B. I. Graubard, and D. Midthune. 1997. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *American journal of epidemiology* 145(1): 72–80.
- Kostka, T., S. E. Berthouze, J.-R. Lacour, and M. Bonnefoy. 2000. The symptomatology of upper respiratory tract infections and exercise in elderly people. *Medicine & Science in Sports & Exercise* 32(1): 46.
- Lagerros, Y. T., L. A. Mucci, R. Bellocco, O. Nyrén, O. Bälter, and K. A. Bälter. 2006. Validity and reliability of self-reported total energy expenditure using a novel instrument. *European journal of epidemiology* 21(3): 227–236.
- Lauderdale, D. S., K. L. Knutson, L. L. Yan, K. Liu, and P. J. Rathouz. 2008. Self-reported and measured sleep duration: how similar are they? *Epidemiology (Cambridge, Mass.)* 19(6): 838–845.

- Laugsand, L. E., L. J. Vatten, J. H. Bjørngaard, K. Hveem, and I. Janszky. 2012. Insomnia and high-sensitivity C-reactive protein: the HUNT study, Norway. *Psychosomatic medicine* 74(5): 543–553.
- Lee, H., I. S. Lee, and R. Choue. 2013. Obesity, inflammation and diet. *Pediatric gastroenterology, hepatology & nutrition* 16(3): 143–152.
- Lee, I.-M., E. J. Shiroma, F. Lobelo, P. Puska, S. N. Blair, P. T. Katzmarzyk, L. P. A. S. W. Group, et al. 2012. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The lancet* 380(9838): 219–229.
- Liu, T.-Z., C. Xu, M. Rota, H. Cai, C. Zhang, M.-J. Shi, R.-X. Yuan, H. Weng, X.-Y. Meng, J. S. Kwong, et al. 2017. Sleep duration and risk of all-cause mortality: a flexible, non-linear, meta-regression of 40 prospective cohort studies. *Sleep medicine reviews* 32: 28–36.
- Liukkonen, T., P. Räsänen, A. Ruokonen, J. Laitinen, J. Jokelainen, M. Leinonen, V. B. Meyer-Rochow, and M. Timonen. 2007. C-reactive protein levels and sleep disturbances: observations based on the Northern Finland 1966 Birth Cohort study. *Psychosomatic medicine* 69(8): 756–761.
- Ludvigsson, J. F., C. Almqvist, A.-K. E. Bonamy, R. Ljung, K. Michaëlsson, M. Neovius, O. Stephansson, and W. Ye. 2016. Registers of the Swedish total population and their use in medical research. *European journal of epidemiology* 31(2): 125–136.
- Ludvigsson, J. F., E. Andersson, A. Ekbom, M. Feychting, J.-L. Kim, C. Reuterwall, M. Heurgren, and P. O. Olausson. 2011. External review and validation of the Swedish national inpatient register. *BMC public health* 11(1): 450.
- Ludvigsson, J. F., P. Otterblad-Olausson, B. U. Pettersson, and A. Ekbom. 2009. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *European journal of epidemiology* 24(11): 659–667.
- Ludvigsson, J. F., P. Svedberg, O. Olén, G. Bruze, and M. Neovius. 2019. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. *European journal of epidemiology* 34(4): 423–437.
- Matthews, C. E., I. S. Ockene, P. S. Freedson, M. C. Rosal, P. A. Merriam, and J. R. Hebert. 2002. Moderate to vigorous physical activity and risk of upper-respiratory tract infection. *Medicine & Science in Sports & Exercise* 34(8): 1242–1248.
- Merk, H., S. Kühlmann-Berenzon, C. Bexelius, S. Sandin, J.-E. Litton, A. Linde, and O. Nyrén. 2013. The validity of self-initiated, event-driven infectious disease reporting in general population cohorts. *PLoS One* 8(4).
- Milner, J. J., and M. A. Beck. 2012. The impact of obesity on the immune response to infection. *Proceedings of the Nutrition Society* 71(2): 298–306.

- Mohren, D. C., G. M. Swaen, P. J. Borm, A. Bast, and J. M. Galama. 2001. Psychological job demands as a risk factor for common cold in a Dutch working population. *Journal of psychosomatic research* 50(1): 21–27.
- Morikawa, Y., K. Kitaoka-Higashiguchi, C. Tanimoto, M. Hayashi, R. Oketani, K. Miura, M. Nishijo, and H. Nakagawa. 2005. A cross-sectional study on the relationship of job stress with natural killer cell activity and natural killer cell subsets among healthy nurses. *Journal of Occupational Health* 47(5): 378–383.
- Mullahy, J. 1986. Specification and testing of some modified count data models. *Journal of econometrics* 33(3): 341–365.
- NCD Risk Factor Collaboration, et al. 2016. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. *The Lancet* 387(10026): 1377–1396.
- Niedhammer, I., I. Bugel, S. Bonenfant, M. Goldberg, and A. Leclerc. 2000. Validity of self-reported weight and height in the French GAZEL cohort. *International journal of obesity* 24(9): 1111–1118.
- Nieman, D., and B. Pedersen. 1999. Exercise and immune function: Recent development. *Sports Medicine* 27: 73–80.
- Nieman, D. C. 1994. Exercise, infection, and immunity. *International journal of sports medicine* 15(S 3): S131–S141.
- Nieman, D. C., D. A. Henson, M. D. Austin, and W. Sha. 2011. Upper respiratory tract infection is reduced in physically fit and active adults. *British Journal of Sports Medicine* 45(12): 987–992.
- Nieman, D. C., and L. M. Wentz. 2019. The compelling link between physical activity and the body's defense system. *Journal of sport and health science* 8(3): 201–217.
- NIH. 1998. Clinical guidelines on the identification, evaluation and treatment of overweight and obesity in adults: The evidence report. Technical report, National Institutes of Health. URL [http://www.nhlbi.nih.gov/guidelines/obesity/ob\\_gdlns.pdf](http://www.nhlbi.nih.gov/guidelines/obesity/ob_gdlns.pdf)
- Nowakowski, S., K. A. Matthews, R. Von Känel, M. H. Hall, and R. C. Thurston. 2018. Sleep characteristics and inflammatory biomarkers among midlife women. *Sleep* 41(5): zsy049.
- Owen, N., G. N. Healy, C. E. Matthews, and D. W. Dunstan. 2010. Too much sitting: the population-health science of sedentary behavior. *Exercise and sport sciences reviews* 38(3): 105.



- Ozga, A.-K., M. Kieser, and G. Rauch. 2018. A systematic comparison of recurrent event models for application to composite endpoints. *BMC medical research methodology* 18(1): 2.
- Padgett, D. A., and R. Glaser. 2003. How stress influences the immune response. *Trends in immunology* 24(8): 444–448.
- Park, S.-G., H.-C. Kim, J.-Y. Min, S. Hwang, Y.-S. Park, and K.-B. Min. 2011. A prospective study of work stressors and the common cold. *Occupational Medicine* 61(1): 53–56.
- Parkin, J., and B. Cohen. 2001. An overview of the immune system. *The Lancet* 357(9270): 1777–1789.
- Patel, S. R., N. T. Ayas, M. R. Malhotra, D. P. White, E. S. Schernhammer, F. E. Speizer, M. J. Stampfer, and F. B. Hu. 2004. A prospective study of sleep duration and mortality risk in women. *Sleep* 27(3): 440–444.
- Patel, S. R., A. Malhotra, X. Gao, F. B. Hu, M. I. Neuman, and W. W. Fawzi. 2012. A prospective study of sleep duration and pneumonia risk in women. *Sleep* 35(1): 97–101.
- Pelfrene, E., P. Vlerick, R. P. Mak, P. De Smet, M. Kornitzer, and G. De Backer. 2001. Scale reliability and validity of the Karasek'Job Demand-Control-Support'model in the Belstress study. *Work & stress* 15(4): 297–313.
- Pizzi, C., B. De Stavola, F. Merletti, R. Bellocco, I. dos Santos Silva, N. Pearce, and L. Richiardi. 2011. Sample selection and validity of exposure–disease association estimates in cohort studies. *Journal of Epidemiology & Community Health* 65(5): 407–411.
- Plymoth, A., M. Rotzén-Östlund, B. Zwegberg-Wirgart, C. Sundin, A. Ploner, O. Nyren, and A. Linde. 2015. Self-sampling for analysis of respiratory viruses in a large-scale epidemiological study in Sweden. *Eurosurveillance* 20(11): 21063.
- Porkka-Heiskanen, T., K.-M. Zitting, and H.-K. Wigren. 2013. Sleep, its regulation and possible mechanisms of sleep disturbances. *Acta physiologica* 208(4): 311–328.
- Prather, A. A., E. S. Epel, B. E. Cohen, T. C. Neylan, and M. A. Whooley. 2013. Gender differences in the prospective associations of self-reported sleep quality with biomarkers of systemic inflammation and coagulation: findings from the Heart and Soul Study. *Journal of psychiatric research* 47(9): 1228–1235.
- Prather, A. A., D. Janicki-Deverts, M. H. Hall, and S. Cohen. 2015a. Behaviorally assessed sleep and susceptibility to the common cold. *Sleep* 38(9): 1353–1359.
- Prather, A. A., and C. W. Leung. 2016. Association of insufficient sleep with respiratory infection among adults in the United States. *JAMA internal medicine* 176(6): 850–852.

- Prather, A. A., N. Vogelzangs, and B. W. Penninx. 2015b. Sleep duration, insomnia, and markers of systemic inflammation: results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of psychiatric research* 60: 95–102.
- Prentice, R. L., B. J. Williams, and A. V. Peterson. 1981. On the regression analysis of multivariate failure time data. *Biometrika* 68(2): 373–379.
- Razonable, R. R. 2011. Antiviral drugs for viruses other than human immunodeficiency virus. In *Mayo Clinic Proceedings*, vol. 86, 1009–1026. Elsevier.
- Reed, R. G., and C. L. Raison. 2016. Stress and the immune system. In *Environmental influences on the immune system*, 97–126. Springer.
- Rencher, A. C., and W. F. Christensen. 2012. *Methods of Multivariate Analysis, Third Edition*. John Wiley & Sons.
- Ritchie, H., and M. Roser. 2020. Obesity. *Our World in Data* URL <https://ourworldindata.org/obesity> (Accessed February 17, 2020).
- Romeo, J., J. Wärnberg, T. Pozo, and A. Marcos. 2010. Physical activity, immunity and infection. *Proceedings of the Nutrition Society* 69(3): 390–399.
- Rothman, K. J., S. Greenland, and T. L. Lash. 2008. *Modern epidemiology*. Lippincott Williams & Wilkins.
- Rowshan Ravan, A., C. Bengtsson, L. Lissner, L. Lapidus, and C. Björkelund. 2010. Thirty-six-year secular trends in sleep duration and sleep satisfaction, and associations with mental stress and socioeconomic factors—results of the Population Study of Women in Gothenburg, Sweden. *Journal of sleep research* 19(3): 496–503.
- Saliba, W., O. Barnett-Griness, and G. Rennert. 2013. The association between obesity and urinary tract infection. *European journal of internal medicine* 24(2): 127–131.
- Sanford, M. T., and L. V. Rodriguez. 2017. The role of environmental stress on lower urinary tract symptoms. *Current opinion in urology* 27(3): 268–273.
- Sanne, B., S. Torp, A. Mykletun, and A. A. Dahl. 2005. The Swedish Demand—Control—Support Questionnaire (DCSQ): Factor structure, item analyses, and internal consistency in a large population. *Scandinavian journal of public health* 33(3): 166–174.
- Sasaki, K., I. Fujita, Y. Hamasaki, and S. Miyazaki. 2002. Differentiating between bacterial and viral infection by measuring both C-reactive protein and 2'-5'-oligoadenylate synthetase as inflammatory markers. *Journal of infection and chemotherapy* 8(1): 76–80.
- SCB. 2018. Living Conditions 1980–2016 – selected indicators. Technical report, Statistics Sweden. URL [https://www.scb.se/contentassets/21dab51c868242c4800f01d58cbfc914/le0101\\_1980i16\\_sm\\_lebr1801.pdf](https://www.scb.se/contentassets/21dab51c868242c4800f01d58cbfc914/le0101_1980i16_sm_lebr1801.pdf)

- Schisterman, E. F., S. R. Cole, and R. W. Platt. 2009. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology (Cambridge, Mass.)* 20(4): 488.
- Scully, D., J. Kremer, M. M. Meade, R. Graham, and K. Dudgeon. 1998. Physical exercise and psychological well being: a critical review. *British journal of sports medicine* 32(2): 111–120.
- Segerstrom, S. C., and G. E. Miller. 2004. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological bulletin* 130(4): 601.
- Shneerson, J. M. 2009. *Sleep medicine: a guide to sleep and its disorders*. John Wiley & Sons.
- Spencer, E. A., P. N. Appleby, G. K. Davey, and T. J. Key. 2002. Validity of self-reported height and weight in 4808 EPIC–Oxford participants. *Public health nutrition* 5(4): 561–565.
- Suarez, E. C. 2008. Self-reported symptoms of sleep disturbance and inflammation, coagulation, insulin resistance and psychosocial distress: evidence for gender disparity. *Brain, behavior, and immunity* 22(6): 960–968.
- Tan, S. Y., and A. Yip. 2018. Hans Selye (1907–1982): Founder of the stress theory. *Singapore medical journal* 59(4): 170.
- Textor, J., B. van der Zander, M. S. Gilthorpe, M. Liśkiewicz, and G. T. Ellison. 2016. Robust causal inference using directed acyclic graphs: the R package ‘dagitty’. *International journal of epidemiology* 45(6): 1887–1894.
- The Obesity Society. 2008. Obesity as a disease: the Obesity Society Council resolution. *Obesity* 16(6): 1151–1151.
- Theorell, T., A. Perski, T. Åkerstedt, F. Sigala, G. Ahlberg-Hultén, J. Svensson, and P. Eneroth. 1988. Changes in job strain in relation to changes in physiological state: a longitudinal study. *Scandinavian journal of work, environment & health* 189–196.
- Theorell-Haglöw, J., E. Lindberg, and C. Janson. 2006. What are the important risk factors for daytime sleepiness and fatigue in women? *Sleep* 29(6): 751–757.
- Trolle Lagerros, Y., E. Hantikainen, D. Mariosa, W. Ye, H.-O. Adami, A. Grotta, F. Ghilotti, and R. Bellocco. 2017. Cohort profile: the Swedish national march cohort. *International journal of epidemiology* 46(3): 795–795e.
- Vgontzas, A. N., E. O. Bixler, H.-M. Lin, P. Prolo, G. Trakada, and G. P. Chrousos. 2005. IL-6 and its circadian secretion in humans. *Neuroimmunomodulation* 12(3): 131–140.
- Vittinghoff, E., D. V. Glidden, S. C. Shiboski, and C. E. McCulloch. 2011. *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*. Springer Science & Business Media.

- Wang, H. E., R. Griffin, S. Judd, N. I. Shapiro, and M. M. Safford. 2013. Obesity and risk of sepsis: A population-based cohort study. *Obesity* 21(12): E762–E769.
- Westbury, L. D., H. E. Syddall, S. J. Simmonds, C. Cooper, and A. A. Sayer. 2016. Identification of risk factors for hospital admission using multiple-failure survival models: a toolkit for researchers. *BMC medical research methodology* 16(1): 46.
- WHO. 2000. Obesity: preventing and managing the global epidemic. Report of a WHO Consultation. Technical report, World Health Organization. URL [http://apps.who.int/iris/bitstream/10665/42330/1/WHO\\_TRS\\_894.pdf](http://apps.who.int/iris/bitstream/10665/42330/1/WHO_TRS_894.pdf)
- WHO. 2010. Global recommendations on physical activity for health. URL [http://whqlibdoc.who.int/publications/2010/9789241599979\\_eng.pdf](http://whqlibdoc.who.int/publications/2010/9789241599979_eng.pdf).
- WHO. 2011. Waist circumference and waist-hip ratio: report of a WHO expert consultation, Geneva, 8-11 December 2008. Technical report, World Health Organization. URL [http://apps.who.int/iris/bitstream/handle/10665/44583/9789241501491\\_eng.pdf](http://apps.who.int/iris/bitstream/handle/10665/44583/9789241501491_eng.pdf)
- WHO. 2013. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Technical report, World Health Organization. URL [https://apps.who.int/iris/bitstream/handle/10665/94384/9789241506236\\_eng.pdf](https://apps.who.int/iris/bitstream/handle/10665/94384/9789241506236_eng.pdf)
- WHO. 2014a. C-reactive protein concentrations as a marker of inflammation or infection for interpreting biomarkers of micronutrient status. Technical report, World Health Organization. URL [https://apps.who.int/iris/bitstream/handle/10665/133708/WHO\\_NMH\\_NHD\\_EPG\\_14.7\\_eng.pdf](https://apps.who.int/iris/bitstream/handle/10665/133708/WHO_NMH_NHD_EPG_14.7_eng.pdf)
- WHO. 2014b. What is Moderate-intensity and Vigorous-intensity Physical Activity? *Global Strategy on Diet, Physical Activity and Health*. URL [https://www.who.int/dietphysicalactivity/physical\\_activity\\_intensity/en/](https://www.who.int/dietphysicalactivity/physical_activity_intensity/en/) (Accessed February 19, 2020).
- WHO. 2018a. World Health Organization. Global action plan on physical activity 2018-2030: more active people for a healthier world URL <http://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf>.
- WHO. 2018b. World Health Organization. Obesity and overweight. Fact Sheet. Updated February 2018. URL <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight1>. (Accessed February 10, 2020).
- Woods, J. A., J. M. Davis, J. A. Smith, and D. C. Nieman. 1999. Exercise and cellular innate immune function. *Medicine & Science in Sports & Exercise* 31(1): 57–66.